语料库翻译学文库

国家社科基金项目成果
上海市十二五规划重点图书

语料库翻译学探索

王克非 著

上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

语料库翻译学文库

语料库翻译学探索

ISBN 978-7-313-08103-2

9 787313 081032 >

定价：41.00元

语料库翻译学文库

国家社科基金项目成果
上海市十二五规划重点图书

语料库翻译学探索

Exploring Corpus-based Translation Studies

王克非 著

上海交通大学出版社

内 容 提 要

本书共分4大部分。第1部分为概论，讲述了语料库翻译学的定义、基本理念、主要研究内容和基本途径，以及发展历程和研究范式。第2部分为语料库与翻译研究之间的关系，内容涉及语料库建设、翻译共性、语言特征宏观与微观描述、翻译文体等方面。第3部分为语料库与翻译教学，对与翻译教学相关的方面进行了描述。第4部分侧重于应用问题的研究，如计算语言学、辅助翻译功能、专门对应语料库等。

本书适用于相关专业的研究人员及教师、研究生使用。

图书在版编目(CIP)数据

语料库翻译学探索/王克非著. —上海：上海交通大学出版社，2012

(语料库翻译学文库)

ISBN 978-7-313-08103-2

Ⅰ. 语… Ⅱ. 王… Ⅲ. 翻译学—研究 Ⅳ. H059

中国版本图书馆CIP数据核字(2012)第007101号

语料库翻译学探索

王克非 著

上海交通大学出版社出版发行

(上海市番禺路951号 邮政编码200030)

电话：64071208 出版人：韩建民

常熟市文化印刷有限公司 印刷 全国新华书店经销

开本：787mm×960mm 1/16 印张：17 字数：309千字

2012年2月第1版 2012年2月第1次印刷

印数：1～3030

ISBN 978-7-313-08103-2/H 定价：41.00元

感谢国家社科基金资助！

语料库翻译学文库
编　委　会

语料库翻译学文库
总　　序

语料库翻译学，即基于语料库的翻译研究；它在研究方法上以语言学理论为指导，以概率和统计为手段，以双语真实语料为对象，对翻译进行历时或共时的研究，代表一种新的研究范式。语料库翻译学一般依靠双语语料库，主要是翻译语料库（translational corpus）、平行/对应语料库（parallel corpus）和类比语料库（comparable corpora），并需分别进行不同的标注。因此对于语料库翻译学的发展来说，有两大关键点：一是双语库的研制，它是技术手段，是基础设施，也体现研制者的研究目的；其加工方式和程度直接影响研究课题的开展。二是研究切入点，即利用双语库能给翻译研究带来什么新发现、新解释，它关系到语料库翻译学如何解决翻译上的各种问题，或如何深化人们对翻译现象和活动的认识。

本丛书为国内（迄今为止也是国际上）第一套基于语料库的翻译研究系列专著，主要选题有“语料库翻译学探索”、“语料库翻译学概论”、“类比语料库及其研究”、“基于语料库的翻译语言特征研究”、“基于语料库的翻译文体研究”、“基于语料库的翻译搭配研究”、“莎士比亚戏剧英汉平行语料库的创建与应用研究”、“语料库与翻译教学”等。其最大特点有三。一是研究者均为国内本领域专家，著作为国家和省部级科研项目成果；二是所收专著各有专攻，覆盖平行语料库、翻译语料库和类比语料库三方面研究；三是成果为第一批与中文相关的语料库翻译学文献，并伴有一批语料库的建立和充实。

语料库翻译学逐步从方法论发展成为连贯、综合、丰富的范式，应用于翻译理论以及翻译的描写和实践等一系列问题的探讨。新的研究范式带来了研究思路的更新和研究重点的转移。编者和作者期盼这套丛书能促进我国的翻译和语料库两方面学术研究的更大发展。

王克非

2011年夏书于北外家中

前　言

在过去的20多年里，语料库语言学的迅速发展给包括翻译学在内的语言研究带来了一场革命。翻译学研究领域于20世纪50年代盛行规范研究，并于80年代开始向描写研究转变。这种研究范式的转变与语料库语言学的兴起与发展不无关系，因为描写翻译学和语料库语言学的主要关注点和研究对象都是实际文本，两者都属于实证研究。

翻译研究从规范向描写范式的转变，加上多语种语料库在90年代的发展，促成了语料库翻译学的诞生。可以说语料库翻译研究已成为目前翻译研究的主流，国内外在过去十年中出版了十多部相关专著与专题论文以及大量的研究论文。然而，这些研究大多可分为两类，一类是介绍语料库在翻译研究中的应用，另一类是用语料库方法来研究具体的语言特征等。

然而，语料库翻译学不仅仅是语料库语言学作为一种研究方法在翻译学中的运用，而是已经发展成为一种复合型的研究范式，为翻译研究、翻译实践、及翻译教学提新的理论框架。虽然语料库语言学和译学界都认识到语料库和翻译研究的有机结合已形成"语料库翻译学"这门新的分支学科，但遗憾的是现有研究还没有对该学科的性质、范围及方法进行准确、系统的学科定位。

王克非先生的《语料库翻译学探索》在这方面作出了有益的探索。作者在系统地回顾语料库翻译研究发展历程的基础上，对语料库翻译学作为一种新的研究范式其理据、总体构架、研究范畴、研究方法及其有待解决的范式问题等学科定位因素进行了详尽的讨论。这本专著的特点可以用三个"兼顾"来概括：兼顾理论研究与实际应用，兼顾理论探讨与实证案例，兼顾翻译研究与翻译教学。这可以说是近年来语料库翻译研究中一部不可多得的里程碑式的力作。

肖忠华

2011年7月于兰开斯特

目　　录

第1部分　概论

第1章　语料库翻译学概要 ………… 3
- 1.1　语料库翻译学的定义 ………… 3
- 1.2　语料库翻译学的基本理念 ………… 4
- 1.3　语料库翻译学的主要研究内容 ………… 5
- 1.4　语料库翻译学的基本途径 ………… 6
- 1.5　小结 ………… 7

第2章　语料库驱动语言与翻译研究 ………… 8
- 2.1　Corpus的含义及语料库的发展 ………… 8
- 2.2　语料库的分类与应用 ………… 10
- 2.3　基于双语库的语言对比分析 ………… 11
- 2.4　基于双语库的翻译研究 ………… 12
- 2.5　研究方法的思考 ………… 15

第3章　语料库翻译学的发展历程 ………… 16
- 3.1　引言 ………… 16
- 3.2　语料库翻译学的主要成就 ………… 16
- 3.3　语料库翻译学带来的启示与思考 ………… 20

第4章　语料库翻译学的研究范式 ………… 22
- 4.1　引言 ………… 22
- 4.2　语料库翻译学的范式理据 ………… 23
- 4.3　语料库翻译学的范式体系 ………… 26
- 4.4　语料库翻译学范式的当下问题 ………… 29

第2部分 语料库与翻译研究

第5章 语料库翻译学的研究课题 …… 35
5.1 引言 …… 35
5.2 国内双语语料库建设与研究 …… 35
5.3 国际双语语料库建设与研究 …… 37
5.4 理论层面的研究课题 …… 38
5.5 应用层面的研究课题 …… 40
5.6 小结 …… 41

第6章 翻译共性研究及其反思 …… 43
6.1 引言 …… 43
6.2 翻译共性的概念 …… 44
6.3 前语料库时期的翻译共性研究 …… 45
6.4 基于语料库的翻译共性研究 …… 47
6.5 翻译共性研究存在的问题 …… 49
6.6 对翻译共性研究的反思 …… 52
6.7 翻译共性研究的未来构想 …… 53

第7章 英译汉语言特征宏观描述 …… 56
7.1 引言 …… 56
7.2 基于语料库的翻译语言特征研究 …… 57
7.3 对英译汉语言特征的宏观描述 …… 58
7.4 结构容量 …… 60
7.5 小结 …… 61

第8章 英译汉语言特征微观描述 …… 62
8.1 引言 …… 62
8.2 英译汉语言的词类分布 …… 62
8.3 英译汉词语的组合方式 …… 66
8.4 小结 …… 70

第9章 翻译特征分析:显化与隐化 …… 71
9.1 翻译的特征与共性 …… 71

9.2 翻译中的显化现象 …… 72
9.3 翻译中的显和隐 …… 73
9.4 广义上的翻译显化现象 …… 77
9.5 小结 …… 78

第10章 翻译汉语的词汇简化与显化 …… 80
10.1 翻译语言的特征与翻译共性 …… 80
10.2 翻译汉语的词汇特征 …… 81
10.3 小结 …… 90

第11章 汉译文学作品的语法显化 …… 91
11.1 翻译共性与语内显化 …… 91
11.2 语料库与翻译小说的"形合度" …… 93
11.3 翻译小说中语法标记的显化 …… 95
11.4 小结 …… 102

第12章 英译汉文本中人称代词的显化特征 …… 104
12.1 引言 …… 104
12.2 英汉翻译中的人称代词 …… 104
12.3 英汉翻译中人称代词主语转换的语料库考察 …… 106
12.4 对翻译中显化的认识 …… 113
12.5 小结 …… 114

第13章 英译汉文学文本中人称代词的变异 …… 115
13.1 引言 …… 115
13.2 人称代词的显化 …… 116
13.3 人称代词的变异:以"他"为例 …… 118
13.4 小结 …… 122

第14章 基于语料库的翻译文体研究 …… 124
14.1 引言 …… 124
14.2 文体的定义及文体学的研究范围 …… 125
14.3 翻译研究的文体学视角 …… 126
14.4 基于语料库的翻译文体学研究框架 …… 129

14.5 小结 …… 130

第3部分 语料库与翻译教学

第15章 语料库翻译教学的理论与原则 …… 133
15.1 引言 …… 133
15.2 对应语料库在翻译课堂教学中应用的理论依据 …… 134
15.3 汉英对应语料库与教学内容的设计 …… 137
15.4 汉英对应语料库运用于课堂上的可能方式 …… 137
15.5 语料库翻译课堂教学对资源的要求 …… 138
15.6 小结 …… 139

第16章 双语对应语料库与翻译教学 …… 140
16.1 引言 …… 140
16.2 双语对应语料库辅助翻译教学及其特点 …… 140
16.3 双语对译检索用于翻译教学 …… 142
16.4 小结 …… 154

第17章 对应语料库翻译教学平台的应用 …… 155
17.1 引言 …… 155
17.2 研究内容 …… 155
17.3 研究设计及步骤 …… 156
17.4 结果与分析 …… 158
17.5 小结 …… 160
17.6 附录:受试报告的分类和统计 …… 162

第18章 从语料库看翻译搭配及其教学 …… 164
18.1 引言 …… 164
18.2 汉译英的搭配问题 …… 165
18.3 基于语料库的汉译英中的搭配教学 …… 167
18.4 小结 …… 176

第4部分 语料库翻译学及相关应用问题

第19章 计算语言学与翻译研究 …… 179

19.1 引言 …… 179
19.2 计算语言学在翻译中的应用 …… 180
19.3 小结 …… 186

第 20 章 汉英对应语料句对检索 …… 187
20.1 引言 …… 187
20.2 语料库的数据库转化设计 …… 188
20.3 用户查询检索接口设计 …… 189
20.4 汉英翻译单位的对应检索 …… 192
20.5 小结 …… 196

第 21 章 语料库辅助翻译功能探讨 …… 197
21.1 引言 …… 197
21.2 机辅翻译工具的类型和特点 …… 197
21.3 双语对应语料库辅助翻译 …… 201
21.4 单语语料库辅助翻译 …… 206
21.5 小结 …… 208

第 22 章 专门对应语料库的研制 …… 209
22.1 引言 …… 209
22.2 现有旅游专门语料库 …… 210
22.3 新型双语旅游语料库的设计思路和特点 …… 211
22.4 小结 …… 218

第 23 章 对应语料库的加工处理 …… 220
23.1 引言 …… 220
23.2 对应语料库 …… 221
23.3 双语对应语料库平台的系统架构 …… 223
23.4 语料对齐加工、校对及入库 …… 224
23.5 小结 …… 230

参考文献 …… 231

后记 …… 248

图 示 目 录

图 2.1　语料库的简单分类(Granger 2003;王克非 2004b:8) …… 11
图 4.1　语料库翻译学体系 …… 27
图 8.1　词类分布折线图 …… 63
图 10.1　原创汉语和翻译汉语文学语料词类频率对比 …… 86
图 10.2　原创汉语和翻译汉语非文学语料词类频率对比 …… 86
图 10.3　各语料库高频词数量比较 …… 89
图 10.4　各语料库词表高频词比例 …… 89
图 11.1　各语料库主要助词频率比较(每百万词) …… 96
图 11.2　翻译和非翻译小说介词频率对比 …… 98
图 11.3　各语料库联合关系连词频率比较 …… 102
图 11.4　各语料库偏正关系连词频率比较 …… 102
图 12.1　英语原文与汉语译文中的人称代词主语 …… 110
图 12.2　英译汉文学与非文学翻译中的人称代词主语转换 …… 111
图 13.1　翻译汉语与原创英汉语人称代词频率比较 …… 117
图 20.1　Web 方式用户查询 …… 191
图 20.2　桌面版用户查询 …… 192
图 21.1　Trados 的检索结果 …… 199
图 21.2　相关检索结果 …… 204
图 21.3　CEPC 检索方式 …… 204
图 21.4　“与其……不如”的索引项 …… 205
图 21.5　“与其……不如”的相关检索结果 …… 206
图 22.1　旅游语料库结构图 …… 214
图 23.1　汉英对应语料库架构设计 …… 223
图 23.2　语料库句子对齐校对界面 …… 225
图 23.3　人工校对语料元数据信息再校对处理描述 …… 227
图 23.4　词类赋码及元信息加工校对 …… 228

表 格 目 录

表 2.1　语料库的三个发展阶段 …… 9
表 2.2　语料库的应用 …… 11
表 6.1　显化的分类 …… 54
表 7.1　原创文本和翻译文本的宏观描述 …… 58
表 7.2　"在……中"的结构容量 …… 61
表 8.1　英译汉词类分布频率 …… 63
表 8.2　指示代词＋数量词组合在两类文本中的分布 …… 67
表 8.3　"是(不)可能(的)"结构在英文源语中的对应项 …… 69
表 10.1　各子语料库的词语变化度比较 …… 82
表 10.2　各子语料库词类频次比较(每百万词) …… 83
表 10.3　各子语料库词汇密度对比 …… 84
表 10.4　各子语料库词表前十位 …… 87
表 11.1　"显化"概念分析 …… 92
表 11.2　各语料库的形合度比较 …… 94
表 11.3　各语料库主要介词词频比较(每百万词) …… 97
表 11.4　各语料库主要连词频率比较(每百万词) …… 100
表 12.1　英汉翻译中人称代词主语考察所选语料 …… 108
表 12.2　汉语非翻译文本参照库所选语料 …… 109
表 12.3　英语原文与其汉语译文中的人称代词主语 …… 109
表 12.4　英译汉文学文本人称代词主语转换类型 …… 110
表 12.5　英译汉非文学文本人称代词主语转换类型 …… 111
表 12.6　英汉文学翻译中人称代词主语的语际隐化 …… 111
表 12.7　英汉非文学翻译中人称代词主语的语际隐化 …… 112
表 12.8　汉语翻译文本与非翻译文本中人称代词主语比较 …… 112
表 13.1　汉英文学(翻译)语料库主要人称代词词频比较 …… 116
表 13.2　各语料库中"他"和人名的复现率 …… 119

表 13.3 多“他”同句的频率 …… 120
表 16.1 汉语中表示条件关系的连词
(陆俭明、马真 1985;黄伯荣、廖序东主编 2002 等) …… 147
表 16.2 英语中表示条件关系的连词(Quirk et al 1985) …… 147
表 16.3 汉英翻译中条件关系句子对应情况(以《围城》英译为例) …… 147
表 17.1 汉语中的动词表达转换为英语 in、at 引导的介词短语
(本科四年级 1 班,共收取报告 24 份) …… 162
表 17.2 汉语中的动词表达转换为英语 in、at 引导的介词短语
(本科四年级 5 班,共收取报告 21 份) …… 162
表 17.3 汉语中的动词表达转换为英语 over、with 引导的介词短语
(本科四年级 2 班,共收取报告 19 份) …… 162
表 17.4 本科四年级 5 班(共收取报告 20 份) …… 163
表 17.5 本科四年级 2 班(共收取报告 17 份) …… 163
表 17.6 研究生一年级(共收取报告 22 份) …… 163
表 18.1 “克服”的中文搭配词 …… 167
表 18.2 “overcome”的常见搭配词 …… 167
表 18.3 “克服”的英语对应词(汉到英方向) …… 172
表 18.4 “克服”的英语对应词(英到汉方向) …… 174
表 18.5 “克服”的英语对应词(双向) …… 174
表 19.1 原文和译文文本特征统计信息 …… 181
表 19.2 汉语原文和英语译文中居前的十个词及其频率 …… 181
表 19.3 原文和译文中居前的十个词类范畴 …… 183
表 19.4 SVD 语句相似度计算结果 …… 185
表 20.1 数据库字段设计及说明 …… 188
表 20.2 利用不同检索字符串确定翻译对应 …… 193
表 20.3 英语译词的选择 …… 193
表 23.1 数据库字段设计及说明 …… 229

第 1 部分

概　论

第1章

语料库翻译学概要

语料库的兴起带来研究工具和研究方法的更新，进而导致语料库语言学和语料库翻译学的产生。本章介绍了语料库翻译学的定义、基本理念、主要研究内容和研究途径，讨论如何在描写研究、应用研究和理论研究三方面展开新课题研究。语料库资源的共享，检索工具的不断研制和更新，使语言和翻译的研究范围从语际对比扩大到语内类比，呈现出双向、多重的对比模式，其意义不仅在理论上，在对翻译现象(也包括双语对比)的描述和翻译教学的改进方面也具有重要的应用价值。

新技术往往带来新的认识工具和新的研究手段。计算机技术的发展使超大规模语料的收集、整理、标注、观察、统计成为可能，从而也使人们对语言的认识有了更宏观的视角，于是发展出现代语料库语言学。这一新型分支学科开展的20多年来，一方面提供了新的语言考察和分析手段，一方面也推进了依据语料库大量语言事实所做的关于语言的理论思考。正如Johansson(2007:1)指出的："通过语料库，我们可以观察先前没有意识到或仅仅隐约觉察到的语言模式"。语料库带来的不只是语言学研究上的更新，还包括翻译学研究方法上的更新。语料库翻译学的发展，使研究者得以对大量真实语料(尤其是双语/多语平行语料，包括书面语和口语)进行对比分析，这为语言对比和翻译研究提供了新的契机。

1.1 语料库翻译学的定义

翻译研究历来重视语言、文本的比较，因此在理论上多借鉴语言学，在方法上多采用描写法、比较法。一脉传承下来，便有了现代的语料库翻译研究。在语料库的支持下，"翻译学理论研究的重心转向，从原文与译文的比较或A语言与B语言的比较，转向文本生成本身与翻译的比较"(Baker 1995:233)。Baker(1996:175)进而提出"基于语料库的翻译研究"(corpus-based translation studies)。此后，Tymoczko(1998:1)将基于语料库的翻译研究称为"语料库翻译研究"(corpus translation studies，简称CTS)。Baker的学生Laviosa(1998:1)更提出翻译研究中的语料库途径是翻译研究的一种"新范式"(a new paradigm)。

鉴于基于语料库的语言学研究通称为语料库语言学，因此，我们也不妨把“基于语料库的翻译研究”称为语料库翻译学。若尝试为语料库翻译学作一定义，我们认为：

语料库翻译学是以语言理论和翻译理论为研究上的指导，以概率和统计为手段，以大规模双语真实语料为对象，采用语内对比与语际对比相结合的方法，对翻译现象进行历时或共时的描写和解释，探索翻译本质的一种翻译学研究方法（参看王克非、黄立波 2007）。

我们说语料库翻译学有语言学理论背景，是因为以 J. R. Firth、M. A. K. Halliday 和 John Sinclair 为代表的英国语言学传统思想是语料库语言学的直接理论源头（Stubbs 1993）。该学派认为语言研究应以真实数据为基础，即以真实文本为主要研究对象进行实证研究，内省式例子本身是解释行为的一部分；将文本整体作为研究的基本单位；基于文本语料的研究以对比为基本模式（同上：8-13）。这些思想一方面打破了传统语言研究建立在内省数据基础上演绎式、规定性的规则系统，另一方面打破了关注语言、忽略言语的倾向，转而对一定社会文化语境中具体的语言使用进行描写和解释，这些间接地为语料库翻译学提供了理论依据。大规模语料库可以帮助翻译研究者重新审视其研究对象，探究翻译研究对象不同于其他研究对象的原因，同时探索指导翻译行为的原则以及制约翻译运作的因素。描写翻译研究摒弃了内省法，从描写真实语料出发，达到对翻译现象的解释和预测。这些均为基于语料库的翻译研究范式提供了直接理论支持。

总之，无论是基于语料库还是为语料库所驱动的研究，都为研究者带来不少新的课题，从而吸引更多的研究人员，并有可能突破原来的研究局面，带来新的启示和进展。

1.2 语料库翻译学的基本理念

20 世纪后半期以来，翻译研究主要还是限于探讨译文与原文之间的关系，这种二元对立的做法常常将译文视为原文的衍生物，在此基础上的理论研究本质上依然是对译文质量的回溯式评估，这种模式在一定程度上局限了翻译研究的进一步发展。从翻译研究本身讲，语料库翻译学有两方面理论发展的支持：其一，语义观转变为情境观，突破了传统的“对等”，将其视为一定社会文化情境中语言使用的对应；第二，描写翻译研究范式打破了原作的主宰地位（Baker 1993：236-241；Laviosa 2002：5）。可见，语料库翻译学的兴起是以突破“对等”为核心概念工具的源文本取向模式（ST-oriented model），由规定转向描写，以大规模翻译文本作为数据来源，结合计算机分析技术的基础上产生的。描写翻译研究率先抛开了传统做

法，对语言相互作用及其影响、目标语文化内部制约翻译活动的种种文本外因素给予了更多关注，将翻译研究视为一个跨学科领域，从多个视角加以审视。

语料库翻译学以翻译本体为研究对象，由大规模翻译文本或翻译语言整体人手，采用语内对比与语际对比相结合的模式，对翻译现象进行描写和解释，探索翻译的本质。

1.3 语料库翻译学的主要研究内容

既然语料库翻译学是基于语料库的研究，语料库翻译学要借鉴语料库语言学的基本方法，包括语料的整理、标注、检索、统计等，但也有它独特之处。首先是所据语料库不同。语料库语言学依据单语语料库即可，语料库翻译学一般要依靠双语语料库，主要是翻译语料库(translational corpus)、对应语料库(parallel corpus)和类比语料库(comparable corpus)。其次是双语语料库在标注方面更加复杂。如翻译语料库需要对翻译、译者等要素加以详细标注，对应语料库需要对两种语料做句子或某种层级上的对齐处理，类比语料库需要对文体、主题、作者、译者等要素加以信息标注。再次是研究对象有别。语料库翻译学探究的是两种语言及其转换的过程、特征和规律。

传统的翻译研究是以原语文本为参照，以忠实程度为取向，主要探讨译文与原文之间的关系或对应关系。语料库翻译学基本上就是语料库语言学加描写翻译研究。这也可以说是一种新的研究范式。因为基于语料库的翻译研究途径正在逐步从方法论发展成为连贯、综合、丰富的范式，应用于翻译理论以及翻译的描写和实践等一系列问题的探讨。新的研究方法会促进研究范式的形成和发展，新的研究范式将带来研究思路的更新和研究重点的转移，这是特别值得注意的。

由此我们可以概括地说，语料库翻译学至少有三大应用的和理论的研究内容：

> 一、大范围的翻译调查，包括翻译教学、翻译文体的考察，以及对应词搭配频率等统计数据的检索与分析；
>
> 二、自动翻译研究，将开展了半个世纪的机器翻译与语料库翻译结合起来，以期取得新的实质性突破；
>
> 三、更广泛更有效地描写翻译研究，包括翻译规范的研究和翻译普遍特征或共性的研究。

Mona Baker 是最早进行语料库翻译学研究的学者之一。她在“语料库语言学和翻译研究：启示与应用(Corpus Linguistics and Translation Studies: Implications

and Applications)"(1993)一文中对这两者的结合作了初步阐发。在大范围翻译调查方面，如翻译文体的考察，她(2000)也率先从语料库角度探讨译者的文体特征，特别是从类符-形符比、平均句长及词项使用特点等方面加以分析。

关于第二个研究指向，主要是将基于规则的研究方法(rule-based approach)同基于语料的研究方法(corpus-based approach)相结合，为自动翻译寻找出更便利可行的途径。

值得关注的是第三点，即语料库基础上的描写翻译研究，特别是有关翻译共性问题的探讨。Baker 在前人研究成果的基础上，提出了翻译普遍特征(universal features of translation)的假设，主要内容是翻译文本中的 1)显化(explicitation)；2)消歧(disambiguation)和简化(simplification)；3)范化(normalization)；4)避免重复；5)凸显目标语语言特征；6)某些特征呈现特定类型的分布。此外，Laviosa 调查了英语翻译文本中的四种核心词汇运用模式；Kenny 通过对原文、译文的语义韵比较，发现译文语言有净化(sanitisation)现象；Øveras 考察了英语-挪威语翻译中衔接层面上的显化现象；王克非据大型对应语料库探讨了译本扩增情况；Laviosa 讨论译文与母语在词汇使用上的不同；柯飞通过语料库考察，发现翻译过程中对原文的模仿可能使译文变得复杂化、冗长化；Xiao & McEnery 发现在"体"标记的使用上，汉语译文比汉语原文多出约一倍；Ebeling 比较了英语和挪威语在存现句使用上的特点；Maia 以双语对应语料库观察英语和葡萄牙语在人称主语使用频率上的差异，等等(王克非 2004b:182-183)。

从这些文献可以看出，语料库翻译学在研究方法上以语言学理论为指导，以概率和统计为手段，以双语真实语料为对象，对翻译进行历时或共时的研究，代表了一种新的研究范式。

1.4 语料库翻译学的基本途径

从古今中外的翻译实践和论述看，最自然的翻译研究途径，也是传统的途径，便是以源文本为参照，以忠实程度为取向，主要探讨译文与原文之间的关系或对应关系。这种二元对立的思路往往将译文视为原文衍生物，在此基础上的理论研究本质上是对译文质量的回溯式评估，这种模式在一定程度上有碍于翻译研究的发展。近 30 年来，国内外学者不断有人尝试打破这种局面，探讨新的研究路径。如 Even-Zohar 等人提出多元系统理论，试图不仅从语言还要从翻译外部即社会文化层面上解释翻译现象，提升了目标语文化语境对于翻译的作用；又如 Holmes 和 Toury 等人提出描写翻译研究法，一方面重视翻译规范(translation norms)的研究，另一方面也试图探究翻译的普遍性特征(universal features of translation)。语

料库的发展，特别是双语语料库的研制，使描写翻译研究得以更充分地展开，基于语料库的翻译研究途径在逐步从方法论发展成为连贯、综合、丰富的范式，应用于翻译理论以及翻译的描写和实践等一系列问题的探讨。

1.5 小结

综上所述，语料库翻译学特别关注三方面的探讨：

第一，描写研究：包括大范围的翻译调查、翻译文体考察，不同译本的参数比较，以及语句的对应情况，对应词和短语的搭配及其频率等统计数据的检索与分析。

第二，应用研究：如自动翻译研究，将开展了半个世纪的机器翻译与语料库翻译结合起来，以期取得新的实质性突破；又如翻译教学，大量的对应本章和语句的呈现，有利于学习者翻译意识的养成和翻译技巧的自我提高；在双语词典的编纂方面，显而易见的是可以丰富例证和提供搭配情况、频率、反向检索等数据。

第三，理论研究：使更广泛、有效的描写翻译研究得以展开，包括翻译规范的研究和翻译普遍特征或曰共性的研究；还可以结合翻译研究进行双语的对比分析，有助于丰富人们对语言的认识以及对语言习得的认识。

第2章

语料库驱动语言与翻译研究

随着计算机技术和语料库语言学的快速发展，大规模语料的收集、整理和标注加工已经实现。如何在此基础上，将描写与实证分析相结合，定性研究与定量研究相结合，更充分地解释各种语言现象，是语料库语言学的重要任务。本章主要讨论在双语语料库驱动下的语言对比分析和翻译研究等方面问题，并探讨语料库研究方法和语料库驱动下的一系列新课题。

2.1 Corpus 的含义及语料库的发展

英文 *corpus* 一词源于拉丁语，本意为 *body*（身体、躯体）。18 世纪后，该词开始用于指称"关于某一主题文字形式的汇编、全集"；到 20 世纪 50 年代，*corpus* 一词逐渐具有现代意义上的"语料库"的含义，指"用于进行语言分析而收集的大量书面语或口语资料"（*OED*，1989：959）；60 年代第一个机读语料库即布朗语料库（the *Brown Corpus*）（Kennedy 1998：23；Kenny 2001：24）在美国建立，标志着现代计算机语料库的诞生，corpus 一词也完成了它到目前为止在意义上的演化①，成为"语料库"。人们通常误以为语料库是伴随计算机的发明而逐步出现的，其实原始意义上的手工语料库出现得更早，甚至可以追溯到中世纪（杨惠中 2002：46）。按照 Kennedy 的说法，最早的涉及语言研究性质的语料库始于 18 世纪以基督教《圣经》为基础的词汇索引，用于语汇检索以及证明《圣经》文本意义在整体上的一致性（Kennedy 1998：13-14）。Francis（1992）认为计算机出现之前的最早的语料库开始于 1775 年 Dr. Johnson 编写的《英语词典》（*A Dictionary of the English Language*）（Kenny 2001：23-24），因为这部词典不仅收词丰富，释义准确，而且大

---

① "corpus"一词在意义上大致经历了这样一个变化过程："身体"→"尸体"→"汇编，全集"→"（口语或书面语的）语料"→"语料库"（*OED* 1989：959）。《牛津高阶英汉双解词典》（第 6 版）中对"corpus"一词的释义为"（书面或口语的）文集，文献；汇编；语料库"（p. 376），由此也可以看出语料库的发展过程。

量引用著名作家及其经典作品中的语言来说明词义和用法。然而，第一个现代意义上的语料库却应当毫无争议地归功于 Randolph Quirk 于 20 世纪 50 年代末在伦敦大学建立的“英语用法调查”(Survey of English Usage)语料库[①](同上：24；王克非 2004b：3)。

语料库从本质上讲，就是“依照某种原则方式所收集的大量文本总汇”(Kenny 2001：22)。今天我们在语言学研究中所谈到的语料库，是指运用计算机技术，按照一定的语言学原则，根据特定的语言研究目的而大规模收集并贮存在计算机中的真实语料，这些语料经过一定程度的标注，便于检索，可应用于描述研究与实证研究。由此可见，以计算机技术的应用为标志，语料库的发展经历了三个阶段：原始语料库、现代语料库和当代语料库。原始意义上的语料库主要指计算机出现之前以语汇索引、词典编纂、方言研究、语言教学研究和语法研究为主要代表的原始手工语料库，其最显著的特点就是对经典文本语料的手工收集。现代意义上的语料库包括运用计算机技术大规模收集多种文本语料的电子语料库，其主要有三个特点：大规模的真实语料；经过一定的标注处理；便于检索。当代语料库是指 90 年代以共建英语语料库(COBUILD)、朗文语料库(LONGMAN)、英国国家语料库(BNC)以及国际英语语料库(ICE)为代表的超级语料库(mega-corpora)，它们的容量更大，收词量大都上亿，覆盖面更宽，应用更广(见表 2.1)。

表 2.1 语料库的三个发展阶段

发展阶段	时 期	特点及应用	典型代表
原始语料库	18 世纪～20 世纪初	手工收集原语料；无标注。主要应用于语汇索引、词典编纂、方言研究、语言教学研究和语法研究	Alexander Cruden 于 1736 年出版的钦定版《圣经》语汇检索；Dr. Johnson 于 1755 年出版的《英语词典》(*A Dictionary of the English Language*)
现代语料库	20 世纪 50～80 年代	由手工收集向计算机语料库过渡；初步依照一定语言学原则进行标注。主要用于语法分析和语言对比研究，逐渐涉及其他领域	Randolph Quirk 的“英语用法调查”(the Survey of English Usage)语料库；布朗语料库(the Brown Corpus)；兰开斯特-奥斯陆-卑尔根(LOB)语料库等

① “英语用法调查”语料库仍以人工收集为主，既包括书面语也包括口语，建立该语料库的初衷是要编写一部“标准”英语语法，起初没有计划借助计算机技术。但其中的口语部分经过计算机处理包括在伦敦-隆德语料库中(Svartvik 1990，转引自 Kenny 2001：24)，因此它是语料库由人工处理向计算机处理过渡阶段的代表。

（续表）

发展阶段	时　期	特点及应用	典型代表
当代语料库	20世纪90年代至今	语料库呈现超大规模趋势；类型更加丰富，应用更加广泛，包括：语言研究、对比研究、翻译研究、教学研究词典编纂，以及机器翻译和软件开发等	共建英语语料库（the Cobuild Bank of English）；朗文语料库（Longman）；英国国家语料库（BNC）；国际英语语料库（ICE）等

需要指出的是，第一，语料库的演变过程说明“语料库”这一概念由来已久，其雏形早在18世纪甚至更早就已经出现，今天语料库语言学中所讲的“语料库”是一个以计算机技术应用为标志的现代概念，也就是说“corpus”或“corpora”并不一定等于现代意义上的“语料库”；第二，早期的语料库主要是对原语料（raw material）的人工收集，用作语汇索引、词典编纂、语法研究、方言研究和语言教学研究的原始语料，以书面语为主，而现代意义上的语料库种类繁多，主要借助计算机手段，在一定的语言学理论的指导下对语料在各个层次上进行了标注，应用范围涉及语言研究、对比研究、翻译研究、教学研究以及词典编纂等众多领域；第三，当代语料库的设计、研制和应用以印欧语系内部语种（尤其是英语）为主，尚未普及到世界各种语言。

2.2 语料库的分类与应用

语料库是一个多角度、多层次的研究工具，因此对其分类也显得纷繁复杂，没有统一的划分。从语料内容上可分为：书面语与口语、共时与历时以及原创语言与翻译语言；从选材方式上可分为：抽样型和监控型；按照语言数量可分为：单语的、双语/平行的和多语的；从用途方面又可分为：普通语料和专门用途语料。而且每个语料库还可进一步细分成一些子库。一般来说，研究目的与方法是语料库类型的根本决定因素，基于此，我们可以对语料库做一分类（见图2.1）。

以上分类主要依据语言数量、文本间是否具有翻译关系或对齐关系（王克非2004b：6-9）。大体上说，语料库的应用涉及以下几个方面：语言研究、对比研究、翻译研究、语言/翻译教学研究，单语/双语词典研编以及机助/机器翻译与翻译软件开发等（见表2.2）。

由表2.2可以看出，与翻译研究有关的语料库主要包括三类：翻译语料库，对应/平行语料库，其中以对应语料库最为突出，其最大的特点在于包括了两种原创语言文本和与之对应的两种翻译语言文本，可以用于多种用途。语言研究和翻译研究向来重视语言材料的收集，而基于大规模电子语料文本库的语言与翻译研究更具有以下四个特点：第一，大规模的真实语料；第二，共时与历时的结合；第三，定

性研究与定量研究的结合；第四，理论与实践的结合。这些特点集中体现了理论研究中的继承性与创造性。

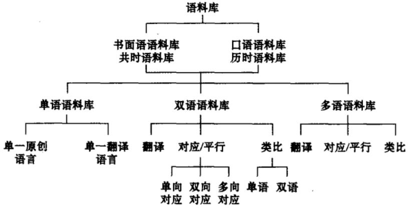


图 2.1　语料库的简单分类

(Granger 2003；王克非 2004b：8)

表 2.2　语料库的应用

语料库类型		语言研究	对比研究	翻译研究	语言/翻译教学	单/双语词典研编	机助/翻译与翻译软件开发
单语语料库	单一原创语言	√				√	
	单一翻译语言	√		√			
双语语料库	翻译	√	√	√	√		
	对应/平行	√	√	√	√	√	√
	类比		√	√			
多语语料库	翻译	√	√	√	√		
	对应/平行	√	√	√	√	√	√
	类比		√	√			

2.3　基于双语库的语言对比分析

对比语言学的源头大约可以追溯到 19 世纪末 20 世纪初的对比分析[①]研究，但

① “对比语言学”(contrastive linguistics)与“对比分析”(contrastive analysis)可互换，王宗炎指出：“对比分析或对比语言学不同于比较语言学(comparative linguistics)。比较语言学是历时性研究，它追究语言之间的谱系关系；对比分析是共时性研究，它追究语言之间的一致性和分歧性—尤其是分歧性。”(2000：111)

最早提出"对比语言学"的通常认为是20世纪美国的语言学家 Benjamin Lee Whorf(王宗炎 2000:112),而后这一学科经历了一个盛衰起伏的过程。对比语言学与翻译研究有着一种不解之缘,两者统一于两种不同语言之间的相互关系当中。两种研究的最终目标不尽相同,但两个领域却存在一种互补关系。对比分析是翻译实践不可或缺的步骤,翻译实践是对比分析的重要手段,在一定程度上两者互为研究方法和对象。对语料的关注是对比语言学和翻译研究最大的共同之处,近年来语料库语言学与计算机技术的结合和飞速发展,使得对大量真实语料(尤其是双语/多语平行语料)的对比分析成为可能,这为对比语言学和翻译研究提供了新的契机。

基于单一语言语料库的语言研究成果非常丰富,基于双语语料库开展的语言对比分析还不多见。这主要是因双语库的建设难度较大,相对滞后。不过随着双语库逐渐为学界所重视,这类研究也开始兴起。其中涉及汉语的双语对比分析有肖忠华、王克非、熊学亮、曹大峰、施建军、柏晓静和詹卫东等(王克非等 2004b),论及语言的体、被动句、把字句、致使结构、动结结构等。国外在印欧语系各语言之间也开展了一些研究。例如 Davidse 和 Heyvaert(2003)的"论英语与荷兰语中的中动结构"(On the Middle Construction in English and Dutch)一文,将语料库和对比研究结合起来,从微观角度针对现有英语与荷兰语中有关中动结构的理论提出了新的假设。作者指出对中动结构的现有分析方法主要有两种:1)作格法(the ergative view),此种观点认为,中动结构中不存在明确的施事者(Agent),因此结构中只有一个参与者,施事行为归于主语,属于唯动格结构的范畴;2)及物法(the transitive approach),该种观点认为中动结构包含两个参与者即隐含的施事者和受事者性质的主语,强调主语的受事者性质。由此可以看出,前一种观点认为中动结构属于包含一个施事主语的主动语态结构,而后者视其为主语为受事者的被动结构。两位作者指出以上两种观点均存在缺陷,提出中动结构本身赋予了主语与动词短语之间一种内在的配价关系,由中动结构所建立的非施事主语与主动动词短语之间的关系从根本说是情态性(modal)的。Hantson(2003)在其"英语动名词结构与挪威语限定词+不定式/at 从句结构"(English Gerund Clauses and Norwegian *det*+Infinitive/at Clause Constructions)一文中从英语-挪威语对比的角度入手,探讨了两种语言特定结构之间的特点及其翻译问题。文中既有对传统观点的维护,也有借助语料库分析方法对两种语言进行相关统计分析和翻译转换问题的探讨。但是这类语言对比分析的研究还不丰富,有待进一步开展。

2.4 基于双语库的翻译研究

纵观翻译理论发展的历史,到20世纪90年代初为止,翻译研究经历了语文研

究、语言学研究、文化研究、哲学研究和认知研究五个范式。早期的翻译理论以文学翻译为主要研究对象，以词汇与语义的关系、是否忠实于原作、译文风格取向以及译者的创造力等语文因素来建立翻译的方法论体系，这一研究范式开创了翻译理论的先河，为后来翻译理论的发展奠定了基础。然而这些理论大多仅限于评论或批评性质的个人经验总结，抽象而模糊。到了 20 世纪 50、60 年代，以 Eugene A. Nida 和 J. C. Catford 为代表，产生了翻译研究的语言学方法体系。他们以语言学理论为指导，采用形式分析的方法从词汇、语法、句法、语义、语用、语篇、功能等层面对原语文本和译语文本的语言进行分析，期望达到对翻译问题的科学解释，这一范式一直延续到今天，依然是翻译研究的主流方法之一。翻译的文化研究范式出现于 80 年代末，以 James Holmes(1972/1988)的"翻译研究的名与实"(The Name and Nature of Translation Studies)一文为起点，Bassnet 和 Lefevere 提出了翻译研究的文化转向，将研究的重心转向了翻译活动所涉及的诸多社会和文化因素。其中以 Even Zohar 的多元系统理论最为突出，改变了以往以"对等"为核心的视角，使翻译研究取得了新的突破。翻译的哲学研究范式由来已久，早在 19 世纪的 Schleiermacher 和 Benjamin 所创立的诠释学(hermeneutic)翻译研究方法就已经为后期的解构主义翻译观奠定了基础。翻译认知研究范式以认知科学的研究为基础，旨在探究翻译过程中译者头脑中的"黑匣子"(black box)；这个领域的研究方兴未艾。以上这五种范式依然渗透于今天的翻译研究当中，使翻译研究呈现一种跨学科的态势。然而翻译研究最为突出的一个问题就是定性研究与定量研究相脱节，理论研究与语言转换的实践相脱节，缺乏客观的量化标准和评估模式。建立基于语料库的翻译研究新方法可以弥补上述不足。

基于语料库的翻译研究方法是建立在以上各种方法论基础之上的，是各种研究方法的整合与延续。它以语言学理论和科学统计方法为指导，以对双语文本转换的真实语料为研究对象，兼及翻译活动中的各种超语言因素，对翻译进行历时与共时的研究，使得翻译理论的研究与语言转换中语料实例重新科学地结合起来。从本质上看，语料库与翻译研究之间存在一定的内在联系。因为翻译研究不外乎涉及文本、文化、译者、翻译过程等语言及超语言因素，而语料库，尤其是双语对应/平行语料库不仅包括了语言信息，也包括了翻译过程涉及的各种超语言因素。

Johansson(2003)认为语料库方法的最大优势在于它可以将理论研究与教学实践真正结合起来。在文末，他对未来的研究方向提出了三点建议：第一，多语语料库研究；第二，翻译语料库和学习者语料库研究；第三，新一代语法书和词典研编。Laviosa(2003)以"语料库与翻译研究"(Corpora and Translation Studies)为题，提出了"基于语料库的翻译研究"(corpus-based translation studies)范式。她从 Toury(1995)提出的"描写翻译研究"(descriptive translation studies)与基于语料

库的翻译研究之间的联系入手，指出了两者的共性：第一，注重从实证的角度对真实语料进行直接观察；第二，对语料库文本的选择并非基于某个固定的定义，而是建立在共识性标准和外部分类的基础上，必要时采取随机抽样的方式；第三，两种方法都坚持认为从实证性研究中得出的一般性结论只有建立在对大量文本语料研究的基础之上才能保证其有效性；第四，适用于研究对象的各种原则都是经过大量系统的研究来发现，并以概率性规则作为表述形式。由此可见，翻译研究的领域尚未最终划定，基于语料库的翻译研究范式是在原有研究范式的基础上孕育产生的，其最大的优势在于其新颖灵活的研究方法和关于不同语言及翻译现象的大量真实语料。

按照 Toury 的观点，翻译理论研究与描述性翻译研究之间是一种互惠的关系，"换句话而言：描述性翻译研究所积累的发现将有助于一系列连贯'法则'的制定，这些'法则'能够说明与翻译活动相关的各个变量之间的内在关系"(1995/2001: 16)。而这样一种互惠的关系是建立在对双语文本的转换进行描述的基础之上。他同时指出纯翻译研究与翻译应用研究之间是一种单向关系，即纯粹的翻译研究只服务于翻译应用研究(同上：17-18)。事实上翻译理论研究、描述性翻译研究和翻译应用研究三者之间存在一种内在的联系：翻译理论提出某种假设或方法论，通过描述性研究对其进行验证，产生完善的方法论体系，然后将其投入具体应用领域。在整个过程中，对语料的实证性分析与描写起着举足轻重的作用。基于语料库的翻译研究是指将原语/翻译文本语料库应用于对翻译产品或过程的实证性研究、对理论建构的阐述以及译员培训等方面的学科分支，它的最大特点就在于借助实证基础上的方法和理论原则，使用归纳和演绎的方法来研究翻译产品和翻译过程。这种新研究方法促进了理论研究、实证研究和应用研究之间的对话(Laviosa 2003;Granger et. al. 2003:45)。因此，基于语料库的翻译研究方法使翻译理论的模式更加科学和完善，这一方法的继承性集中体现在其与描述性翻译研究的共性方面：第一，注重从实证的角度对真实语料进行直接观察；第二，对语料库文本的选择并非基于某个固定的定义，而是建立在共识性标准和外部分类的基础上，必要时采取随机抽样的方式；第三，两种方法都坚持认为从实证性研究中得出的一般性结论只有建立在对大量文本语料研究的基础之上才能保证其有效性；第四，适用于研究对象的各种原则都是经过大量系统的研究来发现，并以概率性规则作为表述形式(同上：49-50)。这样的研究方法具有三个优势：第一，建立不同语言之间在语义层次上的真实对应(这里的"对应"是一个宽泛的概念)；第二，对先前的翻译理论和假设进行重新验证；第三，提出新的假设并加以验证。

Olohan(2003)指出语料库语言学的研究对象不是语言能力(linguistic competence)而是语言应用(linguistic performance)，所以基于语料库的翻译研究就是以文本语

料为基础，从直觉出发产生关于翻译现象的假设，并对这些假设进行系统的研究。她继而提出了翻译研究中语料库研究方法的理论框架：第一，描写研究，显然语料库的方法是从真实的文本语料出发，是对翻译中语言运用的描写；第二，普遍性研究，基于语料库的研究方法以认同/原型的翻译概念为基本概念工具，以多元系统理论（the Polysystem Theory）为指导研究翻译活动中的普遍现象；第三，翻译的语境化研究，即借助翻译文本来重新构建特定社会文化环境中的各种规范和限制性因素。作者同时指出，尽管语料库的研究方法涉及大量的量化数据，但不应过分强调这一方法的"客观性"或"科学性"，因为基于语料库的翻译研究方法同样离不开主观的定性分析。

2.5 研究方法的思考

以上着重讨论了双语语料库的发展、功用以及在语言对比、翻译研究等方面的应用研究。将语料库应用于翻译研究的优势是不言而喻的，但语料库也不是万能的。因此，我们在利用语料库的同时，也要注意其不完善之处。

第一，文本语料不可能穷尽。基于语料库的语言分析和翻译研究主要是建立在对真实语料直接观察基础之上的归纳性研究，那么多大规模的语料库才能称得上是具有代表性？才能使所得出的结论更具有普遍性的特征？

第二，语料库（尤其是对应语料库）的选材过程通常是以建库者的母语语言系统为出发点（Granger et. al.：47），那么选材的标准难免会受到本族语价值体系的干扰，使得对文本的选择带有一定的主观性和偶然性，应当采取什么样的标准来最大限度地摆脱这种影响？

第三，原创语言作品以及翻译作品的质量应当以什么样的标准来衡量？

第四，双语文本的对应问题。上文已经提到对应/平行语料库相对而言与翻译研究的关系最为紧密，双语文本间的对应除了词级、句级、段级或篇章级等形式对应之外，语义对应方式的研究是否应当作为此项翻译研究的重点呢？因为对应语料库的最大特征就在于双语的平行对应，使我们对两种不同语言之间的"一一对应"有了重新的认识，这一点正好应验了雅柯布森（1959）的观点："语言之间真正的区别不在于说话者可以或者不可以表达什么，而在于说话者必须或者决不能表达什么"（钱军等译 2001：266）。

第五，目前世界上印欧语系之外语言（如汉语、日语、朝鲜语、马来语、泰语等）的基于双语库的对比分析和翻译研究成果都还比较少，有待于我们努力。拿汉语来说，英汉两种语言同为世界上的两大语种，分属不同的语系，基于英汉/汉英平行语料库的研究一定会深化我们对许多语言问题的认识。

第3章

语料库翻译学的发展历程

方法论的突破、合理的假设和求证、研究的融通，这三者是语料库翻译学发展的主要脉络。本章简要回顾近15年来新兴的语料库翻译学的发展历程，重点从翻译共性、翻译文体、翻译过程、应用研究（包括翻译教学）等方面评论语料库翻译学的新视点和新阐释。

3.1 引言

语料库应用于翻译研究历史不长。20多年前，Vanderauwera(1985)以五部荷兰语小说的英译本为语料，讨论了翻译语言表现出的语际简化、显化和范化，初步具备了语料库翻译共性研究的雏形。Gellerstam(1986)对比了瑞典语翻译和非翻译小说文本，从语内类比角度讨论了"翻译体"(translationese)的语言特征。这些研究抛开了翻译研究中传统的"对等"理念，开始关注翻译转换中语言之间的相互作用带来的语言变化以及促成此类变化的语言内、外因素，开辟了翻译研究新视角。同时这一方法也被应用于语言对比等研究领域。1993年Baker发表了"语料库语言学与翻译研究：启示与应用"一文，她在文中提出：利用大型的原文加译文的语料库，加上语料库研究方法的探讨，将使翻译研究者得以揭示翻译文本作为沟通媒介的本质(1993:243)。此文可视为"基于语料库的翻译研究"或曰"语料库翻译学"(Corpus-based Translation Studies)研究途径的开通。15年来，各国学者进行了大量的理论与实证探讨，使这一研究途径逐步发展为一种新研究范式（王克非2006)。本章勾勒15年来这一领域的发展，在此基础上讨论该领域的前景以及带给未来翻译研究的启示。

3.2 语料库翻译学的主要成就

语料库翻译学的开展以双语的计算机语料库为前提。20世纪90年代以来，各种双语语料库在各国建立起来。双语语料库主要有翻译语料库(translational

corpus)、对应语料库(parallel corpus)和类比语料库(comparable corpus)三种,各有其特点和用途(王克非等 2004b)。因此对于语料库翻译学的发展来说,有两大关键点:一是双语库的研制,它是技术手段,是基础设施,也体现研制者的研究目的;其加工程度影响研究课题的选择。二是研究切入点,即利用双语库能给翻译研究带来什么新发现新解释或提出什么新问题。限于篇幅,本章主要谈后一点,即语料库翻译学的研究课题。

3.2.1 翻译共性研究(translation universal)

Chesterman(2004a:33)指出,"一门科学只有通过在单个个案之间寻找其相似性并从中加以归纳,才能达到对未来研究或尚未研究个案的预测,从而取得进步。"翻译共性即语料库翻译学试图探寻的翻译语言整体特征。

翻译共性是指翻译语言作为一种客观存在的语言变体,相对于源语语言或目标语原创语言在整体上表现出来的一些规律性语言特征。Chesterman(2004a, 2004b)将其划分为源语型共性(S-universals)和目标语型共性(T-universals)两大类,前者是基于源语文本与译文之间的语际对比关系,关注译者对源语文本的处理方式;后者是关于目标语中翻译文本与非翻译文本之间的语内类比关系,重点放在译者对目标语语言的处理方式上。Baker(1993:243)所谓的"翻译普遍特征"(universal features of translation)即属后一类,她将其定义为"翻译文本而不是原话语中出现的典型语言特征,并且这些特征不是特定语言系统干扰的结果"。柯飞(2005)所提出的"翻译共性"则属前者,是指"译文中呈现的有别于原文的一些典型的、跨语言的、有一定普遍性的特征"。近年来对翻译共性的讨论包括显化(explicitation)、隐化(implicitation)、简化(simplification)、范化(normalization/conventionalization)、集中化(convergence)、整齐化(levelling out)、非典型搭配(untypical collocations)、干扰(interference)、独特项(unique items)等,方法上则是从具体语言因素入手,采用语际、语内或两者结合的对比模式,以概率和统计为手段,尝试从语言之间相互作用在语言层面的表现探索翻译语言的规律性特征,并从语言内、外加以解释。

根据 Baker(1998a),翻译共性研究的目的在于识别翻译文本的典型语言特征,这样不仅可以揭示"第三类语码"的本质,更重要的是能认识影响翻译行为以及隐藏于此类独特语言形式之下的各种具体的制约因素、压力和动机,从而达到对翻译现象的深层解释。不同语种中现有关于翻译共性的实证研究结果不尽一致,既有对现有假设的证明,也有质疑。吴昂、黄立波(2006:300)曾指出,产生质疑的原因有两个:一是对研究对象的界定,二是对各种变量的关注与控制。以译者因素为例,一方面可就不同译者对同一原作或同一作者作品的翻译加以考察,这种研究属于译者文体或翻译文体学范畴。另一方面则是译入还是译出的问题,即将外语文

本翻译成母语,还是将母语译成其他语言,同时涉及翻译方向因素。目前,英译汉文本大多属于译入,即由以汉语为母语的译者完成,而汉译英文本中的情况可以分为汉语母语译者、汉英合作译者和英语母语译者三类,后两类所占比例较小,这也是影响语料平衡性的一个因素。对各类翻译文本中表现出的语言共性的探究有助于我们进一步认识翻译的本质。Toury(2004)提出,对翻译共性的研究既不应过分具体化,也不可层次过高,而应当采用概率思维方式对其加以认识,并以条件式表述方法加以描写和解释;翻译共性研究并非一个存在与否的问题,而是一个解释力问题,即如何借助各种概念工具对翻译现象进行更好的解释。翻译共性研究的最终目标是探究各类翻译活动的一般规律,具体语对(language pairs)中的共性研究都是一般共性研究的组成部分。就此 Eskola(2004)做了一个非常形象的类比,她说翻译共性研究就像一个拼图游戏,有关每一块拼板的使用信息都是完成整个游戏所需信息的一个部分。同样,每一个案研究都会提供特定语对和文本类型中关于共性的局部信息,这些对认识总体共性都是绝对必要的。

3.2.2 译者风格研究(translator's style)

对翻译文体的讨论早已有之,Venuti(1995)曾用"隐身"(invisibility)一词来描述当代英美文化中译者的行为方式及由此而造成的社会地位,他认为译者为迎合读者一味采取归化式的"流畅"(fluency)译法,使译文呈现"透明"(transparency)状态。Hermans(1996:27)提出,翻译叙事话语(translated narrative discourse)在文本中表现出的声音不止一种,其话语呈现(discursive presence)方式也不止一种,他将叙事话语中包含的"第二位"声音称为译者的声音(translator's voice),以此作为译者话语呈现的表征。Hermans(2002:11)指出,译作当中译者异质声音的痕迹随处可循,译作绝非是对原作的透明再现,而是具有多重、分散、混杂和复调(polyphony)性质的双重文本,旁注、脚注、括号说明以及序言等准文本形式无不表明两种不协调的声音共存于译作中,因此以"透明"来评价译作有些过于简单。这些研究主要是从翻译策略及翻译的社会表现形式探讨译者特有的存在方式,是从语言外部来观察翻译整体表现出的特征。语料库翻译研究模式则是将译者以语言和非语言形式所表现出的风格均作为研究对象,扩大了翻译文体的范围。

Baker(2000:245)将文体定义为"留在文本中的一系列语言及非语言的个性特征",翻译文体包括译者的选材、惯用的具体策略,包括前言、后记、脚注、文中词汇注释等,尤为重要的是个性表达方式,即典型的语言运用习惯。由此可见,Baker所谓的翻译文体是一个宽泛的概念,而且这些特征是译者在翻译过程中的无意识产物,并且在一定时间范围内保持相对稳定。Baker 利用类符-形符比、平均句长以及词项 *say* 的使用,对比了 Peter Bush 与 Peter Clark 两位译者的译文后得出结

论：既然从理论上识别某个文学译者的文体是可能的，就有可能借助此类描写来描绘该译者选择再现的世界，帮助研究者将语言习惯与译者的社会文化处境(包括文化之间的关系、预期读者等)结合起来。Olohan(2004：153-167)通过"个体译者的缩略词模式调查"和"应用关键词分析探究译者的词汇选择"两个个案介绍了译者文体研究的具体操作。Olohan 同时指出，影响译者文体的因素包括文化中介的需要、读者期望、意识形态以及编辑与修改等。Malmkjær(2004)以 H. W. Dulcken 对 Hans Christian Andersen 作品的英译为例对译者文体作了进一步探索，初步提出了"翻译文体学"(translation stylistics)这一概念，主要关注在源语文本既定条件下，译者为什么会以特定的方式来塑造译文，对这些现象的解释除考虑语言因素外，亦可从语言外因素，如翻译规范、目标语文本的目的等入手。但在此基础上的大规模研究并未展开，研究所选用语料规模较小，而且仅限于文学翻译。

3.2.3 翻译过程研究

翻译过程是语料库翻译学关注的另一方面。翻译过程研究的对象即译者大脑(常被称为"黑匣子")中的转换活动。由于对此种活动无法直接进行观察，一直以来研究者仅凭借源语与目标语文本的对比，以推测的方式归纳译者在翻译过程中采取的特定策略。20 世纪 80 年代以来，有声思维法(think-aloud protocol)的应用使翻译过程的研究进入一个新的阶段。有声思维法作为一种文本数据采集方法本身就具有语料库研究的雏形，即透过真实的翻译文本来研究翻译过程。Shlesinger(1998)探讨了语料库辅助的口译研究作为基于语料库的翻译研究分支的可行性，将口译研究纳入语料库翻译研究的视野。Englund-Dimitrova(2005)借助有声思维法和计算机记录书写过程等手段，对俄语-瑞典语翻译中转折连词在译文中显化与否进行了考察。此研究的独特之处在于将记录翻译过程的元文本(meta-text)与翻译文本本身结合起来考察翻译过程。传统翻译过程研究的局限在于仅从文本中推测翻译过程，而且语料规模小，语料类型单一，大规模翻译过程语料库的建立可以帮助翻译研究者更为直观的分析以元文本形式表现出的翻译过程。

3.2.4 翻译应用研究

双语对应语料库和类比语料库的应用研究主要是在译者培训、翻译教学方面以及相关软件的开发。代表性的研究包括 Zanettin(1998)、Bowker(1998)、Monzó(2003)、Pearson(2003)等。Bernardini 曾指出，将平行语料检索应用于翻译教学可以培养翻译专业学生的翻译"意识"、"反射"和"应变"(王克非等 2004b)。就学生翻译意识的培养，Chesterman(1997：157-158)有三个建议：第一，为学生提供各种类型的翻译文本，将源语文本与翻译文本加以比较，来认识源语—目标语的各种关

系;第二,将某一源语文本以不同的方式进行翻译,来认识源语—目标语关系在具体实践中的变化;第三,对翻译个案中不恰当的源语—目标语关系进行研究,尝试提出新策略。这些方法旨在从翻译产品入手,通过观摩、对比、分析、借鉴的方式发挥学生的主动性,在评估他人译作和自我实践基础上提高学生的翻译技能,翻译语料库正是实现这些设想的较佳工具。Zanettin(1998)展示了如何将小型普通或专门双语语料库应用于学生自主课堂教学,以加强学生对源语文本的理解,提高他们目标语表达能力。Bowker(1998)选取了两组翻译文本,一组翻译采用传统的练习方法,即以词典等工具书为参考,另一组以一个专门单语语料库作为辅助手段,两者对比表明:后者在理解主题、选择术语、习语表达等方面明显优于前者。Monzó(2003)指出,让翻译学习者认识翻译的最好办法就是让他们置身翻译其中,借助双语对应或类比语料,不仅能让学生直接观察源语语言系统的规则,也可看到翻译文本自身的特征,建立对翻译转换的感性认识。王克非(2007)通过双语对应语料呈现实验,证明学生能够根据自己的观察,归纳和总结翻译技巧,及时评估或反思自己的翻译行为,有助于翻译技巧的掌握。

从形式上讲,应用型语料库主要分对应语料库和类比语料库两大类。前者主要关注语际转换中的翻译现象,还可用来提取双语词汇,供学习者借鉴他人的翻译策略;后者中的目标语非翻译文本发挥着参照功能,供学习者查找常用搭配,分辨特定文本类型中的常规表达方式,为解决特定问题的办法提供依据;从语料内容上应用型语料库可分为普通型、专门型(提取术语和双语词汇)和学习型三类。普通型语料库适用于各类翻译翻译者,专门型语料库,如法律、经贸等可为专业翻译人员提供帮助,学习型翻译语料库以学习者自己的翻译文本为语料,可用于翻译错误分析等方面。Olohan(2004)指出,还可利用网络作为翻译实践中的参照库,在线语料库(online corpus)及翻译记忆系统(translation memory systems)均可充当对应语料库的角色。语料库翻译途径具有很多优势,但同时也有一定局限。Olohan 特别指出,不能因为语料库的方便、快捷、信息量大等特点就过分夸大其在实践与应用方面的作用,因为语料库毕竟是静态的文本,其中受时间限制的模式化策略会抹杀译者的创造性。

3.3 语料库翻译学带来的启示与思考

纵观语料库翻译学 15 年的发展历程,可获得如下一些启示:

第一,一种新研究范式的建立往往以方法论和学科基质的建立为基础,研究方法的改进必定会促进研究范式的发展与更替。语料库翻译学一改先前翻译研究只注重特定语言单位转换的传统,从整体考察大规模文本,从微观入手描写双语转换

的宏观特征，并从语言、认知、社会文化等角度加以解释。可以说，语料库翻译学的研究方法是各种传统翻译研究方法的整合与延续，也是建立新研究范式的尝试。语料库翻译研究本身是一种研究途径，其中涉及的假设、发现、方法与原则都属于方法论的范畴，加上研究者团体及系统的思想阐述，就成为一种研究范式。Lavisoa(1998a;2002:22)指出，正是众多研究者的多样化实证研究与理论阐述使得语料库翻译研究迅速发展，逐步由一种方法论发展成为一种连贯、综合、丰富的范式，应用于对翻译理论、描写以及实践等一系列问题的探讨。

第二，善于提出合理假设，才能开辟新研究课题。建立在直觉基础上的假设是系统研究的出发点。实证研究以假设为出发点，以数据统计为基础，数据是研究对象的一部分，数据以外的因素构成了研究对象的另一部分，前者属于客观的表象，后者较为隐蔽、主观化。Laviosa(2002:2)指出，语料库翻译研究的本质就是对假设进行检验，具体表现为以下步骤：1)提出假设；2)建立研究目标并对假设进行验证；3)描写和分析数据；4)对发现进行理论阐述；5)将假设精确化；6)在此基础上为将来的研究提出新的假设，在重复运作的基础上对理论进行修正和完善并建立初步的方法论体系。以翻译共性研究为例，首先是通过语料库对翻译语言模式进行描写和归纳，然后从语言内部以及跨语言的角度分析各种翻译模式出现的原因，在此基础上，从语言外部，如认知、社会等角度探究各种翻译模式的根源。

第三，翻译研究中的"融通"(consilience)趋势。翻译的本质是本族语在与异域语言交流与相互作用过程中的"扬弃"，这一过程既有碰撞与冲突，也有相互协调与妥协。对翻译本质的认识是一个不断深入的过程，认识的途径既有语言内的也有语言外的，而总的趋势是"融通"①，即"各种知识的统一"。Chesterman(2005:25)指出，从融通的观点看，现代翻译研究声称自己是尝试以跨界的方式探寻对文本、社会和文化之间相互关系的深入认识，达到融通的一条途径就是将翻译研究纳入模因学(memetics)，因为模因(memes)是文化尤其是文化转换的因素。作为实证研究，语料库翻译研究是在描写基础上解释和预测，不仅涉及语料库语言学和翻译研究本身，还涉及社会、文化、认知等各种语言外途径，各种途径的融通是翻译研究的必然趋势。

---

① Chesterman的这一概念来自E. O. Wilson. 1998. *Consilience: The Unity of Knowledge*. New York: Knopf.

第4章

语料库翻译学的研究范式

本章从区别性特征的角度论证了语料库翻译学是翻译学的一种新的研究范式，而不是其他范式的分支或仅仅是一种方法。在此基础上，提出了语料库翻译学的范式体系，包括三个研究层面：理论研究、描写研究与应用研究。本章还讨论了语料库翻译学范式的当下所面临的两个问题：比较中立项问题与逆证的方法论问题。

4.1 引言

1995年，翻译研究中语料库路向的主要推动者 Mona Baker 提出"翻译学理论研究的重心需要转向，即从原文与译文的比较或 A 语言与 B 语言的比较转向文本生成本身与翻译的比较"(Baker 1995:233)。这可以看作翻译研究中语料库路向的正式宣言。1996年，Baker(1996:175)又提出了"基于语料库的翻译研究"(corpus-based translation studies)。1998年，Baker 的学生 Laviosa(1998:1)正式将翻译研究中的语料库路向认定为翻译研究的"一种新范式"(a new paradigm)。同年，Tymoczko(1998:1)将翻译研究中的语料库路向正式命名为"语料库翻译研究"(corpus translation studies，简称 CTS)。在国内，翻译研究中的语料库路向也逐渐得到了认可和肯定。2006年，王克非首次正式提出了"语料库翻译学"，并也将其确定为翻译学的新研究范式(2006:8)。这说明，经过十多年的发展，许多学者认为翻译研究中的语料库路向已取得了质的突破，成为了翻译学的新生长点。同时，也有学者提出了不同的观点。刘康龙、穆雷(2006:61)认为语料库是描写翻译学的一个分支。Olohan(2004:3)认为翻译研究中语料库的使用应看作是一种方法，而不是一种范式。本章将从理论上讨论语料库翻译学的定位，是翻译学的一种新的方法还是一种新的研究范式。我们认为，这是翻译研究中语料库路向的重大理论问题。此外还将探讨，如果翻译研究中的语料库路向是翻译学的新范式，那么其范式体系又是什么？

4.2 语料库翻译学的范式理据

经过80年代的文化转向和90年代的解构主义之后，翻译学可谓陷入了一种危机之中。一方面在本体论上面临着被瓦解的风险，消弭在文化研究之中（吕俊2002）。另一方面在方法论上难以推陈出新，未形成新的研究领域和概念范畴。我们认为，翻译学的学科变化首先来自于翻译思想的变化。刘宓庆（2005：285）认为，当代西方译论思想出现了六种倾向，从而表现出与西方传统译论的疏离：1）在语义内容与表达形式这一对矛盾中，明显向表达形式倾斜；2）在交流内容与交流目的这一对矛盾中，明显向交流目的倾斜；3）在交流意向与交流效果这一对矛盾中，明显向交流效果倾斜；4）在异化与归化这一对矛盾中，明显向归化倾斜；5）在作者功能与译者功能这一对矛盾中，明显向译者倾斜；6）在原文取向与译文取向这一对矛盾中，明显向译文的流利倾斜。由此可以看出，当代西方翻译思想的倾向是形式、目的、效果、归化、译者与译文。也就是说更加重视翻译的经验性研究，即翻译本身的研究，而不寻求与原文的比较。这也是描写翻译学（DTS）的思想源头。

其次，具体的翻译研究路向也发生了变化。Baker（1993）认为，翻译研究中语料库路向的出现主要有三方面的原因：一是新的视角的出现导致了原文地位的削弱。70年代末兴起的多元系统理论对源语文本的地位产生了深远的影响。该派宣称翻译只属于一个文本系统：目的语文本系统。二是20世纪70年代末开始，以原文为导向的对等概念逐渐让位于以目的语系统为起点的概念。翻译研究中源语文本与目的语文本之间关系的语义观的衰落导致了翻译研究中新的意义观出现。意义从概念转向了情景，进而转向了使用，从而推动了翻译研究中的描写倾向，特别是语料库研究。以目的语系统为起点的概念中最重要的就是Toury（1978，1980）提出的规范概念。规范是个描述分析的范畴。规范概念出现后，翻译研究的首要对象不再是单个的翻译文本，而是转向了连贯的翻译文本语料库。三是描写翻译学的兴起。70年代开始，就有学者提出了对翻译研究中的内省法的不满，如Holmes（1972/1988：101）、Newman（1980：64）。他们认为出路就是去观察翻译的实例，然后在此基础上进行归纳与概括。但我们认为，描写翻译学在研究范围上过于注重文学翻译，很难对占实际翻译中比例最大的实用翻译做出合理的解释，其在研究目标上提出的翻译学“定律”也遭到质疑（Tymoczko1998：653；Hermans 1999：92；Venuti 1995：361-363），认为其过于规约性。同时，又未提出具体可行的研究方法，所以难以发展为真正的“普通翻译学”。在此历史背景下，90年代中期以来，语料库翻译学迅速发展起来。语料库翻译学影响了翻译学的方法论基础和研究对象，从而实现了对翻译研究领域的重建。Tymoczko（1998：652-653）指出：“语料库

翻译研究出现在翻译研究发展的关键时刻——整个翻译研究中的各种论争都在语料库翻译研究中得到了反映”。

语料库翻译学借鉴了语料库语言学和翻译学中其他范式的成果，但这并不意味着语料库翻译学仅仅是语料库语言学和翻译学中其他范式的延伸。我们认为，语料库翻译学有自己的特殊的学科地位和区别性特征。下面就讨论一下语料库翻译学与和它关系最密切的语料库语言学、翻译研究中传统的语言学范式、翻译的描写范式之间的区别。

首先，语料库翻译学不同于语料库语言学。语料库翻译学借鉴了语料库语言学的基本理念和基本的研究方法。语料库语言学的重要发现之一就是语言使用中语法和词汇其实是不能分开的，必须整合描写。语料库翻译学不赞同将词汇与语法分开的做法。语料库翻译学使用的工具基本是依赖于语料库语言学的研究工具和软件。但同时，语料库翻译学又有自身的特点。王克非(2006:8-9)认为，主要区别为：一是所据语料库不同。语料库语言学依据单语语料库即可，语料库翻译学一般要依靠双语语料库，主要是翻译语料库、对应语料库和类比语料库。二是标注上双语语料库更加复杂。翻译语料库需对翻译、译者等要素详细标注，对应语料库需做对齐处理，类比语料库需对文体、主题、作者、译者等要素详细标注；这就意味着语料库翻译学的研究范围可以扩展到文本外的因素。三是研究对象有别。语料库翻译学探究的是两种语言及其转换的过程、特征和规律。语料库语言学不注重语言间的转换，而是注重所研究的语言本身。Baker(1996:185)认为，在一定意义上来说，翻译研究颠倒了语料库语言学。翻译研究者最终感兴趣的不是词和句法结构本身。他们感兴趣的是抽象的全局概念，如独立于具体语言并有各种表现形式的显化和简化。我们认为还有一点重要区别是对翻译文本的不同态度。语料库语言学往往不认可翻译文本，许多语言学语料库将翻译文本排斥在外。语料库翻译学认为翻译文本记录真实的交际事件，既不劣于也不优于其他交际事件，具有独立的价值(Baker 1993:234)。

其次，语料库翻译学不同于翻译研究中传统的语言学范式。我们认为，主要有如下区别：一是导向不同。翻译研究中传统的语言学范式以原文为导向，强调与原文的对比，目标是对等。Catford(1965:20)就认为翻译是一项对语言进行操作的工作，即用一种语言的对等的文本材料来替代另一种语言中的文本材料。语料库翻译学以目的语文本为导向，强调目的语中翻译文本与非翻译文本的对比。二是研究重心与语言观不同。翻译研究中传统的语言学范式的重心是单个文本翻译的微观转换分析，分词汇、语法两分。语料库翻译学的重心是多个文本的整体研究，并且不分词法、句法、语义和语用，强调纵聚合。三是理论、概念及范畴不同。翻译研究中传统的语言学范式以语言学理论为基础进行对翻译产品和翻译过程的分

析，其运作在传统语言学框架内。语料库翻译学以语料库的技术手段为基础提出了新的翻译理论、翻译概念和翻译范畴：如简略化、显化、规范化、整齐化等；以及新的技术手段：标注、检索、提取、对齐、分布、频率等。

再者，语料库翻译学也不同于翻译研究中90年代后兴起的描写范式。Laviosa（2003：50）认为描写翻译学与语料库翻译学有如下区别：一是在语料库翻译学范式下，理论、数据、描写与方法体系之间不存在清晰的界限。在描写翻译学范式下，理论、数据、描写与方法体系是四个不同的概念。理论有独立的功能。二是描写翻译学范式允许各种各样的分析方法，只要这些方法是经验性的、严格的和描写的。所以，方法体系只是描写翻译学范式的一个维度，不是其定义整体中的一部分。在语料库翻译学范式下，研究对象就是由方法体系本身产生的，也就是说，方法体系是语料库翻译学定义整体中的一部分。三是在语料库翻译学范式下，唯一的有效数据是真实的口语和书面语文本。针对译者"黑匣子"的任何形式的实验研究都不在语料库翻译学范式的研究范围之内。在描写翻译学范式下，这种实验方法受到鼓励。四是在描写翻译学范式下，翻译学的最终目标是在系统累积事实和局部理论的基础上建立起一个普遍理论。在语料库翻译学范式下，局部理论也是有效的，因为它们反映了各语言对的异质性及其文化特殊性。我们赞成Laviosa的第二种和第三种提法，但不赞成第一种和第四种提法。语料库方法中最重要的步骤不是数据的提取，而是基于当前数据的理论归纳，数据并不直接产生理论。任何学科范式下的理论、数据和描写都是分开的，当然它们之间会相互支撑。语料库翻译学的目标绝不仅仅是满足于局部的理论。语料库翻译学的首要目标就是瞄准翻译的共性，强调翻译文本本身的特质。描写翻译学的目标是翻译活动中的"定律"（Toury 1995）。但在数据的选择方面、研究方法的选择和理论目标的选择上的重大差异已表明语料库翻译学不同于翻译研究中的描写范式。

此外，范式的检验通常通过两个特征。一是它们的成就空前地吸引了一批坚定的拥护者，使他们脱离科学活动的其他竞争模式；二是这些成就又足以无限制地为重新组成的一批实践者留下有待解决的种种问题。凡是共有这两个特征的成就，Kuhn（1962：10）称之为"范式"。取得范式后，科学共同体的收获之一就是找到了该研究领域选择研究问题的标准。只有这些问题是科学的，其他问题都是玄学的。我们认为，语料库翻译学已吸引了一批坚定的拥护者，国外如Baker、Laviosa、Kenny、Olohan、Bowker、Halverson、Puurtinen、Zanettin、Øverås等，国内如王克非等，提出了自己有待解决的问题并做出了自己的贡献。就像语料库语言学发现了"熟语原则"（Sinclair 1991/1999）和"型式"（Hunston & Francis 2000）一样，语料库翻译学发现了翻译普遍性或称为翻译共性（Baker 1993）与第三种代码（Frawley 1984）；提出了自己的研究概念和工具，如频率、搭配、词语索引、分布、多因素分

析等。

在认识论上，语料库翻译学认为翻译具有内隐性与经验性，即翻译的本质不易被察觉，但可以体验。在本体论上，语料库翻译学预设翻译的同质观，并认为翻译的原型未定，所以应采取现实中的翻译观，而不是绝对意义上的规约的理想的翻译观。在方法论上，语料库翻译学不重视直觉和内省方法的使用，重视语料库驱动，利用科学的工具对自然发生的真实翻译的观察及其典型性的研究。杨惠中(2004：编者按)认为，语料库语言学的基本特征是以真实使用中的语言事实为研究对象，采用概率法对语言事实进行分析。语料库翻译学的基本特征是以同质翻译观为基本理念，以真实使用中的大规模翻译事实表现为研究对象，采用概率和统计手段对翻译事实进行分析。根据王克非(2006：9)的分析，我们认为语料库翻译学可定义为：在翻译的同质观观照下，以真实、自然、大规模的翻译语料为研究对象，以概率和统计为方法，定量和定性的描写翻译的特征，理论目标是探索研究和提出翻译的描写、解释与预测规律和机制的一门翻译学科。语料库翻译学范式地位的确立将有利于该范式的全面发展和翻译学整体上的提升。但这并不是说语料库翻译学范式优于其他翻译学范式。对于同一种现象和事实，用何种范式来研究都可能是可取的，但是它们看待研究对象的方式和视角不同。范式决定了我们如何看待对象，把对象看成什么，以及怎样去看对象，并且进而选定自己的研究问题。

4.3 语料库翻译学的范式体系

语料库翻译学的体系可以从两个层面考虑。一是横向上的与翻译有关的语料库建设；二是纵向上的与语料库有关的翻译研究。与翻译有关的语料库可根据语料库中语种的数量、语料库的功能、语料库的方向性、语料库的时间性进行划分。根据语料库中语种的数量，翻译中的语料库可分为单语、双语和多语。根据语料库的功能，翻译中的语料库可分为对应语料库、类比语料库和翻译语料库。在对应语料库中，一般都包括源语文本和目标语文本，根据方向性，又可分为单向的和双向的；根据时间，又可分为共时的和历时的。类比语料库一般是同一语言的，包含原创文本和翻译文本。翻译语料库专门由翻译文本构成，如英国的 TEC 翻译语料库，使用时一般与参照语料库对比使用，如 BNC。Shledinger(1998)还讨论了口译中的语料库问题。目前，口译研究中尚无建成的语料库。口译语料库应包括三类文本：口译文本、相似情景下的原生口头话语和这些文本的书面翻译。语料库的建设是语料库翻译学的前提条件，该项工作目前还有待大力发展。

根据 Holmes(1972/1988)、Toury(1995/2001：10)的翻译学总体框架，纵向上我们提出如图 4.1 所示的语料库翻译学的体系。

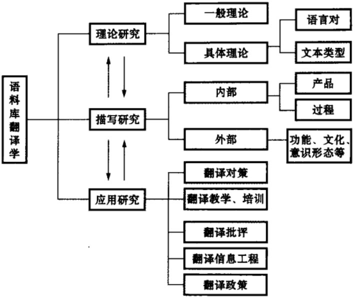


图 4.1　语料库翻译学体系

该体系的特点是把语料库翻译学既看作是翻译学的一个横向分支，又看作是一个纵向分支。总体上，语料库翻译学分为三个大的层面：理论研究、描写研究与应用研究。理论研究分为一般理论研究和具体理论研究。具体理论研究又可分为基于不同语言对的语料库的理论研究（如英汉、法日等）和基于语料库中不同文本类型的理论研究。谭载喜（2005）就提出了比较译学的概念，这是建立普通翻译理论的有效途径。一般理论研究指的是以具体理论研究为基础的翻译的共性和普遍原则的研究。

此外，一般理论研究还可分为产品普遍性和过程普遍性的研究。对于翻译理论的研究模型，Chesterman（2000）认为有三种：比较模型、过程模型和因果模型。比较模型指的就是文本间的比较，是翻译研究中应用时间最长、应用频率最高的模型。到目前为止，大多数的翻译研究都属此类。过程模型强调的是时间概念和通过有声思维等进行数据诱导的实验方法。比较模型的主要手段是寻找两者之间的"对比中立项"（tertium comparationis）。因果模型的形式是：如果 X 出现，Y 就会发生。因果模型强调文本、社会、认知领域的翻译的因、过研究。在这三种模型中，因果模型是最富成效的，可对四种假设进行验证：阐释性假设、描写性假设、解释性假设和预测性假设。描写性假设的形式是：现象 X 的所有实例都具有 Y 特征。解释性假设的形式是：待解释物的致因/原因是 X，或者 E（可能）是由条件 ABC 导致的/影响的。预测性假设的形式是：因素 X 将导致事件或状态 Y；或者说在 ABC 条

件下,事件或状态Y将/倾向发生。理论的最重要的功能就是解释和预测。比较模型可以验证阐释性假设和描写性假设,但不能验证解释性假设和预测性假设,即比较模型不能解释翻译为何是这样,而不是那样,也不能告诉我们这样或那样的后果。所以,语料库翻译学的理论模型应包含两个层面:第一层面是比较模型,第二层面是因果模型。目前大多数的语料库翻译研究属于比较模型下的研究,即探究翻译现象阶段(刘康龙、穆雷 2006:63)。所以,未来的语料库翻译学的理论研究需向因果模型提升。

对于翻译学的理论目标,Toury(1995/2001:264-265)认为应该是找寻翻译"定律",而此定律具有概率性,其形式为:如果X,则Y的可能性就更大/更小。对于复杂的翻译现象,很难用一个定律来概括。徐盛桓(2007:3)认为,当代语言学正经历着原子论向整体论、构成论向生成论、原则论向建模论的转换。我们认为翻译学研究的因果模型也是整体论、生成论与建模论倾向的。Tymoczko(2005)就探讨了作为表征的翻译。对于翻译学的建模,我们认为可有向外的情景建模与向内的心理建模。在心理建模方面,以常规关系、相似、相邻关系为基础,以推导、推衍为手段,徐盛桓(2007:5)提出了话语的"自主-依存分析框架"和"显性表述-隐性表述推导模式"。这可以为语料库翻译学理论建模提供启示。以语料库为基础,以计算机为手段,语料库翻译学正是翻译学理论建模的上佳选择。Holmes(1988:71)、Toury(1995/2001:9)和刘宓庆(2005:15)都认为翻译学是一门经验性科学。所以,对于翻译学来说,描写尤为重要。

对于描写研究,语料库语言学家 Aarts(1999:6-7)认为语言使用的描写模型应满足四点要求:一是模型应允许对数据定性与定量描写的结合;二是模型必须建立语言系统内部的现象与语言系统外部的现象间的关系;三是模型应能允许对各种变体的描写;四是模型应能允许对词、句和篇的整合描写。在对语料进行内部描写的同时,也要对语料进行外部描写。语料库翻译学的描写研究也应该满足这四个方面的要求。Tymoczko(1998:6-7)在描述语料库翻译学未来发展时指出了两个重要方面:一是翻译研究中语言与文化路向的整合;二是意识形态如何影响翻译。所以,语料库翻译学应充分发挥标注丰富的优点,进行语言内外部的整合描写,而目前基本没有进行基于语料库的外部描写。对于应用研究,目前主要是翻译教学和培训。Bowker(1998)的实验表明语料库辅助能够提高翻译质量,特别是在术语和专业领域知识方面。秦洪武、王克非(2007)探讨了对应语料库在翻译教学中的原则与理据。由于理论研究和描写研究的不充分以及方法和工具的不到位,在其他方面尚未展开。对于理论研究、描写研究和应用研究三个大层面之间的关系,我们认为描写研究与理论研究之间的关系是双向的,描写是建立理论的基础,理论又是描写的指导;另一方面描写研究与应用研究之间的关系也是双向的,描写为应用

提供指导，应用又促进描写。但理论与应用之间并不直接关系，而是通过描写的中介。

4.4 语料库翻译学范式的当下问题

虽然，语料库翻译学已取得了不少成就，但也有着自己尚未解决的范式问题。

一是“对比中立项”问题。Baker(2004:7)指出，翻译研究中基于语料库的原始动因就是翻译文本与非翻译文本的相似性与差异性的比较，从而找出翻译文本整体上所特有的型式。Tymoczko(1998:7)认为，语料库翻译学的重要影响之一就是提高了翻译研究中差异(与原生文本)的地位。所以，语料库翻译学中最重要的语料库是类比语料库。也就是说，目前语料库翻译学的主要模型就是比较模型。比较模型的难点就是对比中立项的确定。翻译通常体现的是非对称关系。在翻译中，我们经常选择一些作家、一些体裁、一些主题，而不选择另一些作家、另一些体裁、另一些主题。Bernardini(2005:9)发现，翻译成英语的意大利作家多数是上流作家，而翻译成意大利语的英语作家则多为通俗作家。此外，Kenny(1998:53)还注意到，有的文学体系中根本就没有那种体裁，而是通过翻译来引入的。在这种情况下，我们又如何去对比?

二是方法论问题。Kenny(2000:93)认为，目前专门适合翻译研究的语料库研究方法还不到位。目前的语料库翻译学的理论研究，即翻译普遍性的研究的方法主要是逆证(abduction)。逆证是由美国哲学家 Peirce 在探讨推理的性质时提出的，由 Anderson(1973)引入语言学研究中来并提出了“逆证创新”的概念。从此，逆证概念在语言学研究中广泛应用，并被认为是语言学理论创新的逻辑基础。Peirce (1931-1966:623)早期提出的逆证形式是：

前提:[普遍规则]
　　　　从这个袋子中取出的所有的豆子都是白色的;
　　　　[待解释的观察到的事实]
　　　　这些豆子是白色的;
结论:
　　　　[找出解释上述事实的规则]
这些豆子是取自这个袋子。

对于 Peirce 的早期定义，Deutscher(2002:474)认为此定义有两大缺陷：一是这种前提与结论的变换是肤浅的，并不是真正的概念飞跃，而只是对一显性特质的

概括；二是该定义过于狭隘，逆证仅是一已知规则与待解释的观察到的事实间的匹配。后来，Peirce(1931-1966:189)就将该定义进行了修改：

前提：观察到令人惊奇的事实 C；
若 A 正确，C 就可能是当然的；
结论：因此有理由相信 A 是正确的。

从此，逆证就成了对已观察到的事实进行假设推理的过程。这种推理经过进一步的发展就成为了今天的"寻找最佳解释的推理"(Deutscher 2002:476)。但由于逆证推理属于扩展推理，不是演绎推理，因此在逻辑上仍是站不住脚的。我们认为，受语言学界的影响，目前的翻译普遍性研究就是基于这样的推理模式。下面，我们以翻译普遍性之一"简化"为例阐明这一推理模式：

前提：观察到反常事实：译者经常用上义词、熟悉的词替代译语中没有等值的词；若有一个规则是正确的'翻译是简化的'，'译者经常用上义词、熟悉的词替代译语中没有等值的词'的事实就是当然的；
结论：因此，我们有理由相信有这样一个规则
(所以，翻译是简化的)。

这里的问题是第二个前提没有依据，也就是说逆证使用了缺省的知识。单调逻辑的形式为：$[(p\rightarrow q)\ \&\ p]\rightarrow q$(熊学亮 2007:106)。单调逻辑的正确前提一定蕴含正确的结论，不会出现例外的情况。但非单调逻辑不一样，因为没有逻辑结构上的严密性，所以具有逻辑不可靠的特点。也就是说我们可以有无限个 A。在这里，我们改动一下熊学亮(同上:92)的例子如下：The man sew the rat eating the corn。如果把这句话当作事实 C，则 A 可能的形式有 A1：荒诞的词序排列；A2：有拼写错误；A3：有特殊意义。我们又如何能够判定哪一个 A 是最合理的呢？Peirce 曾提出了简单性、经济性与易证伪的标准(转引自 Deutscher 2002:476)。但这些标准都是比较模糊的。此外，Deutscher(同上:476)认为，Peirce 也没有阐明"假设的可能性"与"理解的量"之间的关系，最可能的解释不一定提供最大量的理解。Hobbs(2004:740)也指出，他没有找到判断这些相互竞争的规则或假设的尺度。这些假设既有可能被验证和认可，也有可能被修正或推翻，并没有逻辑上的必然性。推理有两种形式：演绎推理和扩展推理。扩展推理又分为归纳和逆证。扩展推理虽有貌似的心理现实性，但无逻辑现实性。翻译普遍性的逆证推理也属于扩展推理，所以没有逻辑上的有效性。但这并不是说推理在理论创建中不重要。相

反，推理位于任何理论创新的核心，只是要正确使用，不能误用。Anderson（1973：775）认为，假设效度的检验就是演绎与归纳的事情，演绎预测规则的结果，归纳将假设与新的观察到的现象与结果匹配。

目前，在语料库翻译学领域，在研究目标和研究内容上相互区别的学者持不同的观点与不同的描述和分析，在理论体系上交杂揉合的成分较多。Tymoczko（1998：6）就认为，语料库翻译学的主要吸引力不是首先去找寻普遍性，而是首先以可操作的形式搞清翻译中涉及的不同语言和文化中的具体的语言现象。在研究方法上，语料库翻译学也尚有未解决的问题。这显露出语料库翻译学这一新兴学科早期不够成熟的痕迹，但并不能否定语料库翻译学作为一个新的翻译学研究范式。在本体论、认识论和方法论方面，语料库翻译学都体现了自己独特的学术特征，因而是翻译学的新的研究范式，并有着自己的范式体系。这就意味着语料库范式可以在翻译学的三个层面：理论层面、描写层面和应用层面全面展开，从而为90年代文化转向后的翻译学“本体论上面临着被瓦解的风险，消弭在文化研究之中”的本体危机打开一条出路，让翻译学实现本体回归。对于翻译学来说，重要的是要回答关于翻译的其他学科没有提出的问题，特别重要的是，关于翻译的要求与其他学科有不同的答案的问题。语料库翻译学就是要认可翻译文本的独特价值，确立翻译文本整体上所特有的特征，而不是将翻译置于其他理论框架之下，如文化研究、交际研究等。

第2部分

语料库与翻译研究

第 5 章

语料库翻译学的研究课题

本章介绍了语料库翻译学的新研究范式所带来的主要研究课题及其进展，包括语料库建设、翻译共性研究、翻译语言特征、译者风格研究、语料库翻译应用和基于语料库的翻译教学研究等方面。这些研究为当代翻译学开辟了新的研究课题，是语料库翻译学目前方兴未艾的主要学术阵地。

5.1 引言

本书第 1 部分第 4 章中回顾了语料库翻译学的发展历程。我们知道，任何基于语料库的研究一定离不开其实证的基础和研究的基本源泉——语料库的建设。第 1 部分第 3 章从语料库的定义、分类和发展等方面讨论语料库建设的理论层面。本章集中介绍国内专门用于翻译研究的语料库建设情况。随后逐一介绍语料库翻译学在理论研究层面和应用研究层面的研究课题：翻译共性、翻译语言特征、译者风格、翻译应用和翻译教学等。

5.2 国内双语语料库建设与研究

对于语料库翻译学来所，设计和研制语料库是一个不可或缺的基础。建成了语料库，才有可能开展新课题研究。近年来，新的语料库不断创建，新的课题也在不断探讨。仅以北京外国语大学大型"汉英对应语料库"来说，在该库基础上已开展了许多研究(王克非等，2004b)；后(据 2008 年 8 月上海世界翻译大会最新资料)又有秦洪武凭借该库做翻译语言分析。他发现，汉语翻译语言与汉语原创语言在语言运用的宏观特征有差异：翻译语在词汇密度上低于原创语，在句长上翻译语则高于原创语。这说明翻译语较原创文本在词语使用上不如汉语原创文本用词丰富，而且句子偏长。关键词对比显示，汉语翻译文本和汉语原创文本在名词使用上的差异最为明显。在介词、连词、代词和方位词使用上，汉语翻译文本在使用的数量和频率上都远远高于汉语翻译文本，但在动词使用上两者没有明显差异。在句

法组织方式方面，翻译语言与原创文本语言的差异主要表现在动词和宾语之间的线性距离上，前者明显高于后者。

黄立波则利用该库探究英汉翻译中人称代词主语的显化问题。他对英汉翻译中人称代词主语在文学和非文学两种文体类型中数量、频次和转换类型三方面进行考察，发现：1）文学与非文学英译汉时，人称代词主语数量和频次均呈减少趋势；2）从转换类型看，人称代词主语转换以对应关系为主，语际显化和隐化均不明显；3）相对非翻译汉语文本，汉语翻译文本类比显化突出。在英译汉过程中，人称代词主语语际转换表现出源语迁移现象；语内类比显化突出。他认为根本原因在于英、汉语对形式手段依赖程度的差异，以及语言相对社会地位和认知规律等因素的作用。

燕山大学新近研制的《红楼梦》中英文语料库也值得注意。这是一个一本多译的翻译语料库。该库以《红楼梦》汉语原本与三种英语译本构成句级对应语料库，可以进行中英对照、英英对照研究和互联网自动链接检索，使相关学者得以充分利用语料库储存数据大、计算机运行速度快、有多种语料库工具可以对大量数据进行分析的诸多优势，多方位多角度地对三个英译本在翻译技巧运用、文体设计、语言选择、文化内容处理等方面开支研究。该项目的主要特点是：1）收集汉语 120 回原文和 120 回的霍克斯、闵福德译本、杨宪益夫妇全译本及 56 回的乔利译本。这一个原本、三个译本构成四个单语语料库，可分可合，既可作为一个汉语语料库对应任何一个英译本的双语对应语料库，又可作为三个译本的单语语内语料库，分别用于中英对照和英英对照研究。2）该语料库做了句级平行对应，并对汉英句子的类型、语类、语域、语态及有无修辞、习语等进行了标注。在 ParaConc 等检索工具的帮助下可自由提取语料库中的汉英句对。各个单语数据库支持 WordSmith、AntConc、EditPlus 等检索工具对文本各种信息和语言特色的检索分析。3）可以对该语料库做统计分析，开展各项研究，具体包括中英文各文本词频排列表、词目表、词目分布表，各译本句目、句长、词语搭配、句子类型等语言特色以及翻译技巧的对比分析等。4）该库建立了网络搜索平台，实现互联网自动链接检索。通过该检索平台可进行基于汉英文本内容和句子属性的检索，同时还支持对译本、章回、句子属性、词频及句长等的选择性检索。

此外，各地学者纷纷建立各具特色的语料库，开展自己感兴趣的研究。仍以不久前举行的世界翻译大会情况看，就有 1）Maeve Olohan 以自己的语料库研究翻译的普遍特征问题；2）胡开宝研制莎士比亚汉译语料库，探讨汉译中的显化问题；3）胡显耀基于自建汉译小说语料库进行翻译小说虚词特征研究；4）Wallace Chen 基于语料库探讨建立英汉翻译显化模式；5）王建新基于语料库研究英汉语中的转折连接词；6）戴光荣基于自建平行语料库基础上的翻译显化研究；7）Willy Vandeweghe & Lidia Rura 则试图建立一个平行语料库作为译者的多功能辅助工

具;8)董娜以自建语料库开展林语堂作品的翻译研究;9)夏云以自建小型语料库进行可比文本分析与应用文体翻译研究,等等。北京外国语大学计划创建一个新的语料库:"原创文学与翻译文学语料库",是一个类比语料库(comparable corpus),拟收录约1000万字的20世纪前半叶原创的小说、散文(含小品文)和翻译小说、散文及其英文原文,以便1)考察现代汉语白话文的早期发展,特别是词汇和句式上的变化;2)考察20世纪前期英语文学翻译的特征和变化;3)分析比较原创文学与翻译文学在语言上的异同及翻译对创作在语言及其他表现形式上的影响。

5.3 国际双语语料库建设与研究

当代翻译研究呈现出一种多元化的趋势,除传统的文学翻译研究方法和兴起于20世纪50、60年代的语言学研究方法外,70年代德国的功能理论和以色列的多元系统论,80年代翻译研究的文化转向和翻译研究学派的操纵论、媒介论,以及90年代以来的语料库翻译研究理论与方法,无不在向世人展示翻译研究作为一门独立的学科体系已经崭露头角。下面结合 Alessandra Riccard 所编文集:*Translation Studies Perspectives on an Emerging Discipline*(2002)和 G. Anderman and M. Rogers 合编的 *Translation Today: Trends and Perspectives*(Clevedon: Multilingual Matters Ltd. 2003)以及 S. Granger, J. Lerot and S. Petch-Tyson(编)*Corpus-based Approaches to Contrastive Linguistics and Translation Studies* (Amsterdam: Rodopi. 2003),介绍国际上双语语料库驱动的翻译研究。

从语料库途径探究翻译的一个显著特点就是关于翻译普遍性的讨论。语料库不仅使我们可以同时观察大量的翻译文本,而且可以观察多种语言之间的翻译现象,从而有可能概括出翻译的普遍特性。

Laviosa 是较早依据双语语料库开展翻译研究的学者,她提出了"基于语料库的翻译研究"(corpus-based translation studies)范式。她从 Toury 提出的"描写翻译研究"(descriptive translation studies)与基于语料库的翻译研究之间的联系入手,指出了两者的共性:第一,注重从实证的角度对真实语料进行直接观察;第二,对语料库文本的选择并非基于某个固定的定义,而是建立在共识性标准和外部分类的基础上,必要时采取随机抽样的方式;第三,两种方法都坚持认为从实证性研究中得出的一般性结论只有建立在对大量文本语料研究的基础之上才能保证其有效性;第四,适用于研究对象的各种原则都是经过大量系统的研究来发现,并以概率性规则作为表述形式(Granger et. al. 2003:45-56)。她还曾在博士论文里详尽地论述了翻译普遍性问题,包括翻译中的简化(simplification)、显化(explicitation)、范化(normalization)等特征。总之,基于语料库的翻译研究范式是在原有研究范式的基础上孕育产生的,其

最大的优势在于新颖灵活的研究方法和采用不同语言及其翻译的大量真实语料。

例如，Tiina Puurtinen 以芬兰语单语类比语料库为工具，对原创芬兰语儿童文学与译自英语的儿童文学作品中的语言进行对比。他发现，翻译文本中非限定性结构的出现频次要高于原创作品；翻译文本中所使用非限定性结构在语义、文体和语用功能上与原创作品对该结构的使用不对等。他由此得出结论(Granger et. al. 2003:141-156)：非限定性结构是芬兰翻译儿童文学中翻译语言的一大特征。然而，这一特征与通常的翻译普遍性假设—简化、显化、范化相违背，因为非限定性结构的使用会降低译文的明晰度，增大词汇密度和信息负载，而且大量的非限定性结构背离了原创儿童文学的语言规范。尽管基于语料库的研究方法的优势不言而喻，但这一研究却为我们提出了一些新问题：第一，印欧语系内部基于语料库的研究的结果是否具有更广泛的普遍性呢？如果有，那么这一普遍性存在的基础是什么呢？如何将这样的共性特征应用于翻译实践中？第二，基于语料库的翻译研究方法的有效性到底有多大？对此有效性的衡量应以何为标准？这些问题的探究将有利于基于语料库的研究方法的进一步完善。

Juliane House 则探讨“翻译中的普遍性与文化特性”(Universality Versus Cultural Specificity in Translation, in Riccardi 2002:92-110)。她回顾了翻译研究中语言学范式和文化范式的历史，尤其是对语言-文化相对论与翻译理论及实践之间的关系。她主张采用以语言学为主导、兼顾语用与跨文化的研究范式，并确认了两个基础性的翻译类型/范畴：“显形翻译”(overt translation)与“隐形翻译”(covert translation)。前者是指在译文中保留原作特色，使其明显不具有译语文化中原创文本的特征；后者是指在译作中掩盖原作的特色，使之完全迎合译语读者的口味，看不出其与译语文化中原创文本的差异。她进而提出“文化过滤器”(cultural filter)的概念，用来指译者在翻译中对文化特性进行补偿的手段。通过纯语言学研究范式与纯文化研究范式的有机结合，House 的观点填补了两者之间的鸿沟。她还指出，英语作为全球通用语在文化过滤中的角色所导致的结果就是“文化普遍主义”(cultural universalism)的出现。

从国内国际上的双语语料库建设及相关研究可以看出，语料库翻译学方兴未艾，给翻译研究带来许多新的视角和新的课题。下面继续从理论层面讨论。

5.4 理论层面的研究课题

5.4.1 翻译共性研究

对翻译共性的研究是语料库翻译学的重要贡献和主要学术创新的源泉之一。

早期，Vanderauwera(1985)曾以五部荷兰语小说的英译本为语料，讨论翻译语言中表现出的简化、显化和范化，可视为基于语料库的翻译共性研究的嚆矢。后有Gellerstam(1986)对比瑞典语翻译和非翻译小说文本，从语内类比角度讨论了"翻译体"(translationese)的语言特征。这些研究都不同于传统的"对等"理念指导下的翻译研究，而关注起翻译转换中语言之间的相互作用带来的语言变化以及促成此类变化的语言内、外因素，开辟了新的翻译研究途径。Baker(1993:243-244)在前人研究成果的基础上，提出了翻译普遍特征(universal features of translation)的假设，其中为学界讨论最多的是"显化"、"简化"和"范化"。

其他共性研究例如，Laviosa调查了英语翻译文本中的四种核心词汇运用模式，讨论译文与母语原创文在词汇使用上的不同；Kenny通过对原文、译文的语义韵比较，发现译文语言有净化(sanitisation)现象；Ǿveras以英译挪和挪译英各20本翻译小说的前50句(总共2 000句)为例，考察了英语-挪威语翻译中衔接层面上的显化现象；王克非根据大型对应语料库探讨了译本扩增情况；柯飞通过语料库考察，发现翻译过程中对原文的模仿可能使译文变得复杂化、冗长化(模仿原作)；Xiao & McEnery发现在"体"标记的使用上，汉语译文比汉语原文多出约一倍；Ebeling比较了英语和挪威语在存现句使用上的特点；Maia以双语对应语料库观察英语和葡萄牙语在人称主语使用频率上的差异，等等(王克非 2004b:182-183)。黄立波(2007)重点探讨翻译中连接成分和人称代词主语的显化和隐化，区分出语际显化和语内显化的概念。前者以汉译英为代表，汉语原文中隐而不显的内在连接和人称代词，在翻译成英语时往往需要显化为外在的连接和人称，呈现跨语言显化，即语际显化；后者以英译汉为代表，英语原文中外显的连接成分和人称代词，在翻译成汉语时通常都会被照搬进译文，与汉语原创相比，这些成分的比重会超出很多，呈现出汉语译文和汉语原文的不同，即语内显化。

基于语料库的翻译共性研究呈现如下特点：1)多以共时语料为研究对象，主要依靠计算机技术分析数据，如平均句子长度、形符/类符比率、词汇密度等手段对词汇多样性、信息负载度等方面进行考察；2)以对普遍性的验证和描写为主，解释较少；3)理论阐述与实证研究相结合；4)国外学者多以单语类比语料库为基础，探究相对于目标语原创文本而言的翻译文本特征。因此，翻译共性研究还有不少未能解决的问题，也尚未取得学界的共识。但基于语料库有助于宏观认识翻译是获得认可的。

5.4.2　翻译文体研究

Baker在"语料库语言学与翻译研究：启示与应用"(1993)一文中对这两者的结合作了初步阐发。在大范围翻译调查方面，如翻译文体的考察，她(2000)也率先

从语料库角度探讨译者的文体特征,特别是从类符-形符比、平均句长及词项使用特点等方面加以分析。她(2000:245)将文体定义为"留在文本中的一系列语言及非语言的个性特征"。她认为翻译文体包括译者的选材、惯用的具体策略,包括前言、后记、脚注、文中词汇注释等,尤为重要的是个性表达方式,即典型的语言运用习惯。她初步分析的结论是,从理论上识别某个文学译者的文体是可能的,因而有可能借助此类描写了解该译者选择再现的世界,有助于研究者将语言习惯与译者的社会文化处境(包括文化间的关系、预期读者等)结合起来,也是将翻译的/语言的研究同社会的/文化的研究结合起来。Baker 的同事 Olohan(2004:153-167)则通过"个体译者的缩略词模式调查"和"应用关键词分析探究译者的词汇选择"两个个案分析,讨论了译者文体研究的具体操作。她指出,影响译者文体的因素包括文化中介的需要、读者期望、意识形态以及编辑与修改等。Malmkjær(2004)以 H. W. Dulcken 对 Hans Christian Andersen 作品的英译为例,对译者文体作了进一步探索,提出了"翻译文体学"(translation stylistics)这一概念。她关注在源语文本既定条件下,译者为什么会以特定的方式来塑造译文,对这些现象的解释除考虑语言因素外,亦可从语言外因素,如翻译规范、目标语文本的目的等入手。这是国外学者的一些探索,而相关的涉及汉语的英译汉或汉译英的翻译文体还未得到研究,也是留待我国学界探讨的课题。

5.5 应用层面的研究课题

语料库的兴起带来研究工具和研究方法的更新,进而导致语料库语言学和语料库翻译学的产生,带来研究范式上的变化和研究课题的丰富。Labov(1975)很早以前就指出,语言学研究范式的每一次重大变化都源于对于该领域内数据看法的改变(转引自 Stubbs 1993:24),这一点也适合我们评价语料库翻译学。语料库资源的共享,检索工具的不断研制和更新,使语言和翻译的研究范围从语际对比扩大到语内类比,呈现出双向、多重的对比模式,其意义不仅体现在理论上,在对翻译现象(也包括双语对比)的描述和翻译教学的改进方面也极有应用价值。

5.5.1 翻译应用研究

基于语料库的翻译应用研究主要是将基于规则的研究方法(rule-based approach)同基于语料的研究方法(corpus-based approach)相结合,为自动翻译寻找出更便利可行的途径。从理论上说,翻译记忆就是一种双语对应语料库,规模足够大,语料足够丰富,标注方便实用,它就是解决自动翻译的极好途径,是一个极有前景的大课题。实际上,目前不少的翻译公司也在尝试以双语对应语料库或储存

的已译语料作为自动翻译的辅助工具。

5.5.2 翻译教学研究与实践

另一个重要的应用领域就是翻译教学或译员培训，现在有不少研究者开始探讨利用双语语料库进行辅助翻译教学和翻译自主学习的探讨。双语对应语料库辅助翻译教学，是指借助大规模原文及其对应译文的电子文本，辅以计算机统计手段的翻译教学模式。这一模式旨在从翻译产品入手，通过观摩、对比、分析、借鉴的方式发挥学生的主动性，培养学生的翻译意识，在评估他人译作和自我实践的基础上提高学生的翻译技能。国际上关于利用双语对应语料库提高翻译教学的研究可参看 Zanettin, F., S. Bernardini & D. Stewart(2003)所编《翻译教育中的语料库》(*Corpora in Translator Education*)论文集，其中不少论文有一定参考价值。Bernardini(1997)指出，翻译教学辅之以对应语料库检索，便于翻译专业学生形成一种翻译的“意识”、“反射”和“应变”，这些技能“使专业译员有别于那些不熟练的业余爱好者”。利用平行对应语料库(配以合适的检索工具)便于查找特定表达方式的译法，使术语和短语的翻译更加准确、地道，而且常常可以提供多个翻译选择或翻译参考，比双语词典的例证更为丰富，更为真实。

至于翻译自主学习，双语对应语料库无疑是很好的工具。Monzó(2003)指出，要让翻译学习者认识翻译，最好的办法就是让他们置身其中(exposure)，而双语库是创造此种翻译环境的一条途径，因为通过双语平行语料，不仅能让学生直接观察原语语言系统的规则，也能看到翻译文本的自身特征，建立对翻译转换的感性认识。王克非、秦洪武、王海霞(2007)的一项基于语料库的翻译自主学习实验也表明，语料库和语料分析工具为发现式学习活动提供了有力的辅助工具，它能激发学生的兴趣，促使学生将注意力集中在意义和形式的关联上，提供轻松的课堂互动环境。从外语学习和翻译技巧觉识上看，学习者通过观察双语对应语料可能对搭配、语义偏向和语义韵等现象更加敏感；翻译技巧和策略可以通过学生之间和师生之间的互动性讨论逐步形成，这有助于学生形成稳定、持久和灵活的翻译策略，而单靠教师课堂传授翻译技巧很难做到这一点。

5.6 小结

由上述可知，语料库翻译学方兴未艾，新课题新探讨层出不穷。语料库的兴起带来研究工具和研究方法的更新，进而导致语料库语言学和语料库翻译学的产生，带来研究范式上的变化和研究课题的丰富。在描写研究、应用研究和理论研究三大方面，语料库翻译学都大有可为。目前国内外学者们纷纷创建和研制各种语料

库，并开展各具特色的研究，大大丰富了学界对翻译的认识，推进了学术的进步。无论是基于语料库还是为语料库所驱动的研究，都为研究者带来不少新的课题，从而吸引更多的研究人员，并有可能突破原来的研究局面，带来新的启示和进展。在这个意义上可以说语料库翻译学代表了一种新的研究范式。

第6章

翻译共性研究及其反思

语料库方法应用于翻译研究推动了对翻译共性的研究，但翻译共性研究早在语料库方法介入之前就已开始。本章追述这一研究的历程及主要研究成果，同时论及目前对此论说的质疑，指出以Mona Baker为代表的基于语料库的翻译共性研究在研究对象的界定和方法论方面的确存在一些问题，提出对翻译共性的研究不应仅仅局限于翻译过程本身，还应当从具体语言对、翻译方向、译者因素、文体类型等多个视角进行。

6.1 引言

人们往往对翻译文本抱有一些偏见，究其根源在于翻译文本中使用的语言，这种特殊的语言变体向来被冠以“翻译体”(translationese)的“恶”名。对翻译语言本身特征的关注可以追溯到20世纪六、七十年代。Even-Zohar(1978)指出，在翻译中“我们可以发现一些语言模式，既不能根据原语言语库，也无法按照目标语言语库来加以解释”(Hatim 2001:152)。Duff(1981)将这种听起来像“外”语的翻译语言视为“第三类语言”(the third language)，Frawley(1984)称其为第三类语码(the third code)。翻译体与第三类语码常被混用，但两者有差别，前者是偏离目标语规范的极端形式(Shuttleworth & Cowie 2004:188)，而后者则是翻译共性研究的数据来源。

20世纪90年代以来，新型翻译语料库所拥有的大规模原文和译文文本语料与相关计算机分析技术结合所产生的数据为翻译研究提供了新视角，基于语料库的翻译研究方法应运而生，经过十多年的理论阐述与实证研究已经发展成为一种新研究范式。其研究范围覆盖了从翻译过程到产品的各种翻译现象，特别是翻译语言普遍特征、翻译过程、翻译转换与规范、译者文体等方面。这些领域中又以对翻译共性的研究最为突出。

6.2 翻译共性的概念

翻译共性(translation universals),亦称翻译普遍特征或翻译共性[①],Baker(1993:243)将其定义为"翻译文本而不是原话语中出现的典型语言特征,并且这些特征不是特定语言系统干扰的结果"[②],即指翻译语言作为一种客观存在的语言变体,相对于目标语原创语言,在整体上表现出来的一些规律特征。这一定义有两重含义:第一,翻译共性是特定语言模式的概率性分布特征,主要在归纳的基础上获得;第二,翻译共性由翻译过程本身造成,与两种语言系统之间的差异无关。

Chesterman(2004a,2004b)将翻译共性划分为两大类:源语型共性(S-universals)和目标语型共性(T-universals),前者是基于源文本与译文之间的关系,关注译者对源语文本的处理方式;后者是关于目标语中翻译文本与原创文本之间的关系,重点放在译者对目标语语言的处理方式上。Baker(1993:243)所谓的"翻译普遍特征"(universal features of translation)即属于后一类,她将其定义为"翻译文本而不是原话语中出现的典型语言特征,并且这些特征不是特定语言系统干扰的结果"。柯飞(2005)所提出的"翻译共性"则属于前者,是指"译文中呈现的有别于原文的一些典型的、跨语言的、有一定共性的特征"。

Baker在前人研究成果(Blum-Kulka 1986;Toury 1980,1991;Vanderauwera 1985;Shlesinger 1991等)的基础上,首先提出了基于语料库的"翻译普遍特征"的假设,包括六个方面:1)译文显化程度显著提高;2)消歧和简化;3)合乎语法性;4)避免重复;5)超额再现目标语语言特征;6)翻译过程往往会导致某些语言特征表现出特定的分布类型(同上:243-245)。随后,在此基础上的翻译共性研究不断深入、细化,逐步形成了一整套方法论。然而,其中一些个案研究的结果却与共性假设或多或少有些出入,由此引发了关于翻译共性存在与否的讨论。2001年在丹麦哥本

---

① 翻译共性指"译文中呈现的有别于原文的一些典型的、跨语言的、有一定普遍性的特征"(柯飞 2005)。

② 按照Baker的原文,翻译普遍特征是指"features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems."(Baker 1993:243)

哈根和芬兰萨翁林纳分别举行的两次会议①上，与会者就这一话题进行了深入探讨。

在对假设进行验证和阐述的过程中，Baker 的翻译普遍特征被归纳为简化(simplification)、显化(explicitation)和范化(normalization)三个方面②。因为基于语料库的翻译共性研究离不开对前人研究成果的传承，我们参照 Laviosa(2002)和 Olohan(2004)的做法，以 Baker(1993:233-250)的“语料库语言学和翻译研究：启示与应用”一文为界，将翻译共性研究划分为前语料库(pre-corpus)和基于语料库(corpus-based)的两个时期加以评述。

6.3 前语料库时期的翻译共性研究

Bernardini & Zanettin(2004)认为基于语料库的翻译共性研究存在的问题主要表现在对共性概念的界定和所采用的方法论两方面。关注的对象不同，所采用的研究模式就会有差异。在翻译共性研究中存在两种对比模式：1)原文文本与译文文本的语际对比模式；2)目标语中译文文本与原创文本的语内类比模式。早期对翻译共性的关注，主要是从语际对比的角度出发，代表性研究有 Vinay & Darbelnet (1958)、Nida (1964, 1969)、Blum-Kulka & Levenston (1983)、Vanderauwera (1985)、Blum-Kulka(1986)、Toury(1980,1991)、Shlesinger(1991)等。

语料库用于翻译研究可以追溯到 80 年代(Laviosa 2002:1,21)，但当时文本采集的规模通常较小，并且多以手工统计为主。这里所谓“前语料库时期”就是指大规模机读翻译文本用于翻译研究之前，通过人工采集原文与译文文本，并对与翻译有关的语言现象进行人为对比、分析和统计的时期。这一时期的翻译共性研究主要表现为从词汇、句法和文体等角度关注简化、显化和范化三方面问题。

翻译共性研究中所谓的简化，是指译者在目标语文本中对原语文本中的“语言/信息/语言和信息下意识地简单化”(Baker 1996:176)处理的倾向。Blum-Kulka &

---

① 两次会议分别为：1)2001 年 8～9 月于丹麦哥本哈根举行的第三届 EST 大会，议程中包括关于普遍性问题的小组讨论，部分论文收录于 Hansen, Gyde, Kirsten Malmkjær and Daniel Gile (eds.). 2004. *Claims, Changes and Challenges in Translation Studies. Selected Contributions for the EST Congress, Copenhagen, 2001*. Amsterdam: John Benjamins Publishing Company; 2)2001 年 10 月于芬兰萨翁林纳(Savonlinna)举行的关于翻译普遍性研究的大会，主要论文收录于 Mauranen, Anna and Pekka Kujamæki (eds.). 2004. *Translation Universals: Do They Exist?* Amsterdam: John Benjamins。

② 对翻译普遍性特征包含的具体内容目前尚未有完全统一的看法，但相关理论阐述与实证性研究主要集中于这三个方面。

Levenston(1983:119)将翻译中的词汇简化(lexical simplification)定义为"用少量的词处理原文的过程和/或结果"。语料库翻译研究途径正式提出之前对简化现象的探讨主要有 Blum-Kulka & Levenston(1983)、Vanderauwera(1985)、Baker(1992)等,内容主要涉及对词汇、句法和文体的简化,又可分为由两种语言系统差异所造成的强制性简化,和由翻译过程本身、译者偏好、目标语语文规范等因素造成的非强制性简化两大类①。Blum-Kulka & Levenston 所提出的五项策略②中,除两项外其他均属前者,与下文基于语料库的简化研究不同。Vanderauwera 的研究可以说具备了语料库翻译共性研究的雏形,她所论及的简化手段大都属于非强制性简化的范畴,但在讨论中句法与文体之间的界限不甚清楚,总体上看只是翻译实践中译者遵循目标语语言使用规范的一些归化手法。

显化的概念最早由 Vinay 和 Darbelnet(1958)作为一种文体翻译技巧提出,是指将原语中隐含的信息在目标语中加以明示,此类信息在原文本中可以根据语境或情境获得(1958/1995:342)。在语料库翻译方法提出之前,有关显化的研究主要有 Vinay & Darbelnet(1958)、Nida(1964,1969)、Vanderauwera(1985)、Blum-Kulka(1986)、Klaudy(1998)等。Vinay 和 Darbelnet 是从对比文体的角度提出这一概念的,他们将显化和隐化视为一组对立的文体翻译技巧,常与信息的获得与损失相联系。所涉及的显化包括词汇显化和信息显化两种,这两类显化表面上都是语境或情境要求的结果,但从本质上讲依然是语言系统差异造成的,属于强制性显化的范畴。Nida(1969:164-5)指出好的译文往往多少会比原文长,主要是因为在翻译过程中,译者会通过明示原文中的含蓄信息在适当限度内增加译文的冗余度。他主要是从语言文化差异和方便译文接受者理解原作的角度来探讨翻译中的新信息现象。Vanderauwera 指出了译者所采用的各种显化技巧,这些技巧均不在语言系统差异影响之列,但与译者的个人文体偏好不无关系,这在一定程度上为后来用语料库研究译者的文体奠定了基础。Blum-Kulka 从翻译中衔接与连贯的转换入手提出了显化假设(the explicitation hypothesis):翻译过程会使译文相对于原文更加冗长,其表现形式为衔接方式的显化程度提高(1986:19)。这一假设有两个特点:第一,将显化视为翻译过程内在的属性,摆脱了语言系统差异的影响,使显化研

---

① Kinga Klaudy 在讨论显化时,将其分为四类:强制性显化、非强制性显化、语用显化和翻译内在显化(Baker 2004:82-83),其中非强制性显化的部分内容和翻译内在显化属于语料库翻译普遍性探讨的范围,此处借用 Klaudy 的说法。

② Laviosa 在其"Universals of translations"(Baker (ed.) 2004:288)一文中似乎将 Blum-Kulka & Levenston(1983)中的简化策略分为六类,原作中实际分为五类(Blum-Kulka & Levenston 1983:126-36)。

究具体化；第二，从句法的非强制性转换入手，将译者的文体偏好作为参数之一。Blum-Kulka的论断为后来基于语料库的翻译共性研究提供了重要思路。Klaudy（1998）将显化分成强制性显化、非强制性显化、语用显化和翻译内在显化四类（Baker 2004:82-83），各类之间尽管存在一定程度的重合，却让我们认识到显化现象的多样性和多层次性，在此基础上能对翻译共性做出更加全面的认识。

范化是指“遵循，甚至于夸大目标语中典型的模式和做法的倾向”（Baker 1996:176-177）。Vanderauwera（1985）和Toury（1995）是前语料库时期范化研究的主要代表。Vanderauwera发现译文语篇在标点、选词、文体、句子结构以及语篇构建方面的转换均趋向于目标语语言传统规范的特点（Laviosa 2002:55）。Toury从描写翻译研究的途径入手提出的标准化（standardization）法则既有语际转换基础上对比，也包括目标语内翻译文本与原创文本的比较（Paloposki 2001），这样的论断拓宽了翻译共性研究的视角，并且为进一步对翻译共性进行解释提供了多重角度。

相对于基于语料库的方法，前语料库时期的翻译共性研究除具有Laviosa（2002:57-58）所提出的五个特点外，还表现出：1）每个研究都是从各自不同视角，如对比语言学、文体学或翻译研究的角度出发，在对个别概念的认识上存在一定的偏差；2）在关注焦点上，词汇、句法、文体等层面均有涉及，但缺乏统一的指导原则；3）对共性的研究主要以语际转换为基础，即相对原语而言译语的普遍特征。尽管如此，这些前期研究为以后基于语料库的翻译共性研究奠定了基础。

6.4 基于语料库的翻译共性研究

自Baker（1993）提出基于语料库的翻译普遍特征研究以来，近十多年围绕这一主题的研究大量涌现，具代表性的主要有Baker（1993，1995，1996）、Laviosa（1997，1998a，1998b，）、Kenny（1998，2001）、Ǿveras（1998）、Olohan & Baker（2000）、Mauranen（2000，2004）、Puurtinen（1998，2003a，2003b）、Tirkkonen-Condit（2000，2002，2004）、Englund-Dimitrova（2003）、Frankenberg-Gacia（2004）、Klaudy（1998，2001，2003）、Kenny（1999，2001，2005）、Johansson（2004）、王克非（2003）、秦洪武和王克非（2004）、贺显斌（2003）、柯飞（2003，2005）等。这些研究的主要内容是对简化、显化和范化等普遍特征在各语种大型双语语料库中的实证检验。这些研究的结果既有对原共性假说的支持，也有质疑，同时还有新的假说提出。

Laviosa（2002:58-59）指出：Baker关于简化的研究有三个特点：第一，翻译共性是翻译过程本身内在的特征；第二，共性研究所采纳的是翻译产品/目标语取向（product-/target-oriented）的视角和基于语料库的方法；第三，以类比语料库

(comparable corpora)作为语料数据来源。这三个方面也是语料库翻译共性研究不同于前语料库时期相关研究的主要所在，Baker 的思想仍是这一研究途径方法论的基础。其中 Laviosa 对简化(1998b)以及核心词汇模式(1998a)的分析将共性研究完全局限在目标语语言环境内，树立了共性研究的单语类比语料库研究模式(monolingual comparable corpus-based model)，尽管这一模式打破了传统上原语语篇的主宰地位和对等的观念，但在客观上也造成了排除原语文本的做法，缺少了对共性现象进行充分解释的一个维度(Kenny 2005；Bernardini 2005)。

Kenny(2005)还特别指出对应语料库的运用并不是要重新回到原语与译语文本二元对立的老问题上去，而是要探究如何在单语类比语料库研究范式的基础上，进一步通过对应语料库来完善翻译普遍现象研究的综合法。Øveras(1998)则以英语-挪威语平行语料库为基础，重新从句法角度在语篇衔接层面上考察显化，从而将前语料库时期的研究途径与基于语料库的方法结合在一起。Olohan & Baker 对非强制性关系词 *that* 在原创英语与翻译英语中使用的研究虽都表明显化是语言转换中的一个普遍倾向(Olohan 2004：93-97)，但其采用的依然是单语类比语料库的研究模式，而且此类语言现象仅局限于英语，因此只是对翻译共性一个间接的证明。Mauranen、Puurtinen 和 Tirkkonen-Condit 等在翻译芬兰语语料库(Corpus of Translated Finnish)基础上对翻译芬兰语的研究结果不仅有对翻译共性的支持，同时也包括对现有翻译共性的质疑和挑战，为这一研究提出了一些新问题。柯飞以汉英对应语料库[①]为基础，分别从宏观和微观角度对显化进行了验证和探讨，并特别指出："作为一种翻译现象，显化(以及隐化)不应只是狭义地指语言衔接形式上的变化，还应包括意义上的显化转换，即在译文中添加了有助于译文读者理解的显化表达，或者说将原文隐含的信息显化于译文中，使意思更明确，逻辑更清楚"(2005：306)。

Englund-Dimitrova(2003)通过有声思维法和书写过程的计算机记录等手段，对俄语-瑞典语翻译中转折连词在译文中显化与否进行了考察，各项数据表明：翻译中的显化与译者的翻译经验有关，职业译者往往对显化策略的使用要比学生译者多。这一研究从社会文化的视角指出，显化是俄-瑞翻译中职业译者所遵循的一种翻译规范。Frankenberg-Gacia(2004)以葡-英双向平行语料库(Compara)为基础，从译文与原文文本长度角度探讨了非强制显化。该研究以定量分析的方法证明了翻译文本中词汇数量的整体增加表明译文往往比原文更为明晰，而且这并不受制于两种语言之间的差异。这些研究，特别是翻译芬兰语语料库和汉英/英汉平

---

① 该语料库由北京外国语大学中国外语教育研究中心研制，具体内容参见：王克非等，2004，《双语对应语料库：研制与应用》。

行语料库基础上的研究，将翻译共性延伸到了印欧语系以外的翻译研究中。

总的说来，基于语料库的共性研究呈现出四个特点：1)大多数研究都是以共时语料为研究对象，主要依靠计算机技术分析数据，如平均句子长度、形符/类符比率、词汇密度等手段对词汇多样性、信息负载度等方面进行考察；2)以对共性的验证和描写为主，解释较少；3)理论阐述与实证研究相结合；4)主要以单语类比语料库为基础，探究相对于目标语原创文本而言的翻译文本特征。这项研究总体上经历了一个由单语类比语料库范式向单、双语结合语料库(包括单语参照库和双语对应库)综合法发展的过程，研究内容由词汇向句法、语篇层次进行延伸，宏观与微观、理论阐述与实证性研究并举。

针对 Olohan & Baker(2000)采用的类比模式，Kenny(2005)以德英文学文本平行语料库对非强制性 *that* 研究的结果重新验证后指出：英语译者在 *say* 之后使用非强制性 *that* 的大多数情况并非为了去对应德语中的 *daβ*，表面上属于语法显化，但原语文本中可能还使用了除 *daβ* 以外的其他连接词形式来明示从句的叙述性质，因而很难绝对地说英语译文中包含 *that* 的做法全都属于显化。Kenny 还特别指出，平行语料库的运用并不是要重新回到原语文本与译语文本二元对立的老问题上去，而是要探究如何在单语类比语料库研究范式的基础上，进一步通过平行语料库来完善翻译共性研究的综合法(同上)。

翻译共性存在一定的层次性和复杂性，研究的首要问题是对研究对象加以明确界定，其次应根据综合的原则，采取将单语语内类比同双语语际对比相结合的方法，提高研究结果的信度。

6.5 翻译共性研究存在的问题

由上述可以发现，翻译共性研究目前依然处于对假设的验证阶段，学者们就现有假设的共性(hypothesized universals)尚未达成共识，对一些潜在共性(potential universals)探索也在进行当中。那么翻译共性果然存在吗？若存在，是以何种形式存在？我们首先看一看对翻译共性的一些质疑，然后对产生问题的原因加以分析。

对翻译共性的质疑主要表现在三方面：第一，对方法论的质疑。面对翻译共性研究中单语类比语料库(monolingual comparable corpus)的传统范式，Kenny 最先将平行语料库引入翻译共性研究中(1998，2001)，并且提出如何在类比语料库研究发现的基础之上开展平行语料库研究，将后者作为前者的补充研究(2005：157)，以弥补翻译现象研究中因忽略原语文本而造成的缺陷。针对 Olohan & Baker(2000)采用的类比语料库方法，Kenny 以一个德英文学文本平行语料库对其研究结果进

行重新验证后指出：英语译者在 *say* 之后使用非强制性 *that* 的大多数情况并非为了去对应德语中的 *daβ*，乍看起来属于语法上的显化，但还需要确认一下原语文本中是否使用了除 *daβ* 以外的其他连接词形式来明示从句的叙述性质，如果有的话，就很难断定英语译文中包含 *that* 的做法全都属于显化的现象（Kenny 2005：161）。由此可以看出，部分基于单语类比语料库研究范式的翻译共性研究结果在方法上有待于平行语料库研究的进一步验证和补充。Bernardini（2005：6）也指出，"正是翻译作为沟通媒介的本质，使得翻译分析中所使用的唯目标语取向的方法在方法论上值得怀疑"，例如类比语料库设计中由于不平衡性所造成的可比性缺乏（lack of comparability）问题。可见，方法论上的缺陷是造成问题的根本原因之一。

第二，一些实证性个案研究的结果与共性假设相背离。Puurtinen（1998，2003a，2003b）以芬兰语类比语料库为基础，从句法角度将芬兰语翻译儿童文学中非限定性结构作为衡量可读性的标准之一，对翻译芬兰语的语言特征以及意识形态规范进行考察。她发现原创芬兰语作品中非限定性结构的出现频次低，而译自英语的翻译作品中，非限定性结构的使用频次明显高于当代原创芬兰语作品，由于非限定性结构的使用降低了译文的明晰度，增大了词汇密度和信息负载，而且大量的非限定性结构背离了芬兰语原创儿童文学的语言规范，这一发现显然与 Baker 等所提出的翻译文本的显化相互矛盾。Laviosa 根据自己的研究（1998a，1998b）也指出，通过英语类比语料库所发现的简化证据在各个文本类型中的表现并不完全统一（2002：77），如翻译文本中平均句子长度在新闻和散文两类文本中呈现出相反的趋势，前者比原创新闻英语短，而后者明显长于原创散文。这样看来，文体类型不仅是简化研究中的一个变量，而且将平均句子长度作为简化的一个参数还有待于进一步验证。Kenny（2001）以德英文学文本平行语料库（GEPCOLT）为基础对翻译中的范化和译者创造性进行考察发现，在翻译中词汇范化现象的确存在，但原语文本中的创造性词汇在大多数情况下却没有发生范化（2001：187、210）。这说明对范化现象的研究还是不能完全脱离原语语篇的因素。Mauranen（2000）对翻译和原创芬兰语学术及非小说文本的考察发现，多词语符列（multi-word strings）在翻译文本中的多样化程度高于其在原创文本中的程度，这就说明在学术文本翻译中的范化特征并不明显，也说明这一普遍特征还会受到具体文体类型的局限（Puurtinen 2003a）。Johansson（2004）针对 Jean-Claude Chevalier（1995）提出的译文通常选择人称类主语（human subject）的倾向是简化或范化的表现这一观点，通过英语-挪威语平行语料库（ENPC）对英译挪中主语选择的调查发现：绝大多数情况下，译文会保留原文的主语形式，主语类型的改变往往由两种语言系统之间在结构或文体上的差异造成。以上这些个案研究的结果或多或少地表现出与现有共性假设的矛盾，暴露出存在的一些问题。

第三，印欧语系之外的共性研究对原有假设的挑战。印欧语系以外的翻译共性研究不仅有对已假设共性的支持，还包括对潜在共性的进一步探索，某些发现还对原有的假设提出了挑战。Tirkkonen-Condit(2000,2002,2004)以芬兰语类比语料库为基础在探究翻译文本与原创文本之间是否存在系统的差异，以及是否由于翻译过程造成了翻译语言与原创语言的差异等问题时，提出了独特项假说(the unique items hypothesis)：目标语语言中的某些语言项/元素在源语中缺少语言对应项，相对于原创文本而言，此类语言项/元素在翻译文本中的频次低。也就是说，如果将目标语原创语言规范视为常态，那么翻译语言特征通常被认为是对常态的偏离(Tirkkonen-Condit 2002:209)，即非标准化或非范化，这一点与"遵循甚至于夸大目标语中典型模式和做法"(Baker 1996:176-177)的范化相背离。柯飞(2003)以英汉/汉英平行语料库为基础对汉语中的特殊句式"把"字句在英汉转换中的分布特点进行考察发现：1)翻译汉语作品中的"把"字句要比原创汉语中的"把"字句频次高；2)文学类作品比非文学类作品中"把"字句的频次高；3)由于"把"字句适于表达复杂和细微的意思，因此把字句的频次说明了翻译文本中显化现象。这一结果又与 Tirkkonen-Condit 的独特项假说形成鲜明对比。柯飞(2005)进一步指出：翻译中的显与隐是共存的，翻译中隐、显现象和程度可能与语言的形式化程度和翻译方向相关，同时还会受到译者、社会文化和文本因素的影响。

产生以上质疑的原因主要有两方面：第一，研究对象的界定。Bernardini & Zanettin(2004)认为基于语料库的翻译共性研究在对共性概念的界定和所采用的方法论两方面存在问题。以 Baker 为代表的学者，将翻译共性的研究对象局限于翻译过程本身所造成的、翻译文本相对于目标语内原创文本而表现出的语言特征，仅以单语类比语料库方法论为模式，使得实证研究中语内类比和语际对比所造成的两大类翻译共性混杂在一起；第二，各个研究中对各种变量的关注与控制不尽一致。具体涉及语言对、翻译方向、译者因素、文体类型等多种因素，如译入文本与译出文本的差别，不同文体类型文本之间的差别都是影响研究结果的重要因素。

基于语料库的翻译研究方法至少有两大特点：可观察性(observability)和可重复性(replicability)。这一研究模式面对的是实实在在的语料，从假设出发，在各种语言中不断加以验证和置疑，周而复始，使得研究逐渐深入。如上所述，翻译共性研究中各种变量的介入，加上翻译共性假设本身在概念界定和方法论方面存在的问题，导致了个别实证性研究对翻译共性的质疑。随着越来越多的大型双语语料库在各语种中的建立，以及实证性研究的进一步深入，需要重新界定研究对象，发展方法论，在更加充分描写的基础上，从语言、认知和社会文化等视角建立一套系统的解释和预测机制，并将研究结果应用于(或验证于)翻译实践和翻译教学中。

6.6 对翻译共性研究的反思

Baker(1993)将翻译共性定义为“翻译文本而非原话语中出现的典型语言特征,并且这些特征不是特定语言系统干扰的结果”。这一定义将翻译共性局限于由翻译过程本身造成的语言特征,排除了语言系统差异的干扰。然而在实际操作中,我们很难将翻译过程和语言系统差异这两个因素造成的共性清楚地区别开来,从而造成对这一概念认识上的差异。其次,Baker(1993)所倡导的类比语料库基础上翻译语言与原创语言的对比研究奠定了翻译共性研究的方法论基础,但在翻译研究中排除原语文本因素的做法本身就值得商榷。Bernardini(2005:6)认为“正是翻译作为沟通媒介的本质,使得翻译分析中所使用的唯目标语取向的方法在方法论上值得怀疑”。甚至 Baker 本人也存在自相矛盾的地方:她一方面指出翻译普遍特征的研究主要是借助类符-形符比率(type-token ratio)和词汇密度(lexical density)等手段,来探究译文文本相对于目标语原创文本而体现出的一些规律性语言特征(1995:235-238),另一方面,在讨论显化、简化和范化时却分别以相对于原文文本的文本长度(text length)、译者对原作标点符号的改变,以及原语文本及语言的相对社会地位等语际对比参数作为讨论的依据(1996:180-184)。

可见,对共性概念的界定和方法论的选择是共性研究的首要问题。Paloposki(2001)在谈及简化时指出:一些学者视简化为语际转换过程中的现象,而另一些学者却将简化研究局限于翻译语言与非翻译语言所构成的单一目标语语言内,这样一来,尽管讨论的都是翻译共性,对象却各不相同,在方法论上缺乏一定的统一性。针对这一认识上的混淆,Chesterman(2004a,2004b)区分了源语型和目标语型两类共性。可见,不同的视角便会产生出对翻译共性不同的分类方法,如绝对的/相对的;语际的/语内的;强制性的/非强制性的;语言差异造成的/翻译过程造成的;译者的有意识行为/无意识行为;广义的/狭义的;语言形式的/语义的。因此在具体研究首先应当对共性的概念加以明确界定。在方法论方面,单语类比语料库研究法存在明显的局限性,具体表现为:1)排除源语文本而孤立的讨论翻译问题;2)目标语内翻译文本与原创文本可比性的缺乏;3)一定程度上忽略了对翻译方向、译者因素、文体类型等各种变量的控制,导致研究结果信度与效度的缺乏。Mauranen & Kujamæki(2004,Introduction)指出:就加深我们对翻译的认识而言,共性研究的价值在于从与共性相关的认知、社会和语言三个主要领域出发发展理论,积累证据,因此没有必要在方法论上固守任何一元论。一些实证性研究的结果(Øveras 1998;Kenny 2001,2005;柯飞 2003,2005 等)也已经证明,基于语料库的共性研究需要打破类比语料库研究模式的垄断,重新将原语文本因素纳入其研究的视野。

除此以外，一些局限因素的存在也是造成翻译共性研究缺乏统一性的原因。具体表现包括：第一，翻译活动内在的不对称性。如 Zanettin(2000)的研究表明，以意大利语出版的图书中翻译作品的数量远远大于原创作品，其中一半译自英语，而以英语出版的图书中翻译作品所占的比例却要小得多，其中译自意大利语的作品还不足 5%。而且两种翻译作品在各自的目标语文学系统中所享有的地位也不相同，译自英语的意大利语作品是意大利流行小说的主宰形式，而译自意大利语的英语作品在英国是一种高雅文学。倘若从翻译作品中选择意-英双向平行语料库的语料，它们就缺乏一定的可比性(Olohan，2004：47)。当然，这里还需要对两种翻译作品分别在两国内的读者群加以考察，本章在此限于篇幅暂不作深究。尽管此类问题主要存在于使用范围较窄的一些小语种当中，但在此类语料库基础上的研究结果就会受到影响；第二，语言对的局限。每种语言中都有自己独特的语言表现形式，如英语中的非强制性 *that*、汉语中的"把"字句、芬兰语中表示充足性的动词 *ehtii*(has enough time/is early enough or quick enough)等，对这些语言项的分析结果只能是对共性假说部分的支持或反对，需要在此基础上做进一步的归纳；第三，文体类型的局限。根据 Laviosa(1998a，1998b)的研究，同样为叙述性文本，在散文和报刊中分别体现出的普遍特征却存在一定的差异。翻译共性在文学文本与学术性、非小说文本中表现也不尽相同。这些都说明共性特征也会受到文类的局限。由此可见，对翻译共性还应当进一步的分门别类，使之具体化，并在各汉语体类型的双语平行语料库的基础上重新加以验证。

6.7 翻译共性研究的未来构想

语料库翻译研究范式产生前后，翻译共性研究在内容上既包括语际转换中译文相对于原文的语言特征，也包括目标语内部译文文本相对于原创文本的语言特征；采用的方法包括对比分析、单语类比语料库模式、双语平行语料库模式等；研究性质以实证性描写为主。由于基于语料库的翻译共性研究最初主要在以英语为主的印欧语系中开展，因而在此基础上所建立的方法论和一些研究成果还有待于进一步在非印欧语系中确认。随着各类翻译语料库(包括翻译语料库、类比语料库、平行语料库)在各语种中的建立，相关研究的结果在许多方面(如 Puurtinen 1998，2003a，2003b；Mauranen 2000；Tirkkonen-Condit 2000，2002，2004；柯飞 2003，2005 等)对原有基于印欧语系的研究发现提出了质疑，这就要求我们重新认识和界定研究对象，在原有方法论的基础上对研究方法加以修订和完善。

具体可以从以下几个方面入手：第一，重新界定研究对象。就翻译共性研究的内容而言，不应仅局限于翻译过程或语内类比的共性，而应将源语型和目标语型两

类共性包括进来加以梳理，首先有针对性地进行局部描写、分析和解释，然后再上升到整体归纳。以汉英/英汉翻译中的显化为例可做出这样的分类(见表6.1)。

表6.1 显化的分类

<table>
<tr><th colspan="3">显化类型</th></tr>
<tr><td rowspan="4">语内对比($T_1 \leftrightarrow T_2$)</td><td rowspan="2">形式上(T_1 相对于 T_2)</td><td>强制性</td></tr>
<tr><td>非强制性</td></tr>
<tr><td rowspan="2">意义上(T_1 相对于 T_2)</td><td>强制性</td></tr>
<tr><td>非强制性</td></tr>
<tr><td rowspan="4">语际对比($C_{st} \leftrightarrow E_{tt}/E_{st} \leftrightarrow C_{tt}$)</td><td rowspan="2">形式上(E_{tt}相对于 C_{st}/C_{tt}相对于 E_{st})</td><td>强制性</td></tr>
<tr><td>非强制性</td></tr>
<tr><td rowspan="2">意义上(E_{tt}相对于 C_{st}/C_{tt}相对于 E_{st})</td><td>强制性</td></tr>
<tr><td>非强制性</td></tr>
</table>

(T_1=翻译汉语；T_2=原创汉语；C_{st}=汉语原文；E_{tt}=英语译文；E_{st}=英语原文；C_{tt}=汉语译文)
(强制性显化=语言系统差异导致的显化；非强制性=翻译过程导致的显化)

首先依据不同的语料库类型将显化分为语内对比和语际对比显化两类，前者是基于目标语中翻译文本与原创文本的对比研究，后者则是原语与译语文本比较分析的结果；任何语言类型通常都包括形式和意义两部分，在此基础上各类进一步分为形式上和意义上的显化[①]两类；然后再依据语言系统差异和翻译过程两个诱因对强制性与非强制性显化[②]进行区分。

第二，在指导原则上，以平行语料库为基础，在翻译研究中重新引入源语文本。翻译活动离不开源语文本，即便是在以目标语为取向的描写翻译研究和基于语料库的翻译研究范式下对翻译共性的探讨也同样如此，翻译文本的独立性并不排斥其与原语文本之间的联系。

第三，在方法论方面，以类比语料库的研究成果为出发点，在平行语料库中重新进行验证、分析和归纳。基于语料库的研究的一大特点就是研究过程的可重复性(replicability)，这也应当是翻译普遍特征研究的一大特征。Kenny(2005:155)认为平行语料库可以接过类比语料库的研究发现，继而尝试在源语文本中寻求对这些发现的解释。

---

① 根据王克非的观点，形式上的显化是狭义上的显化，指“语言衔接形式上的变化”；意义显化是指“在译文中添加了有助于译文读者理解的显化表达，或者说将原文隐含的信息显化于译文中，使意思更明确，逻辑更清楚”(柯飞 2005)。

② 这里借用 Klaudy 的说法，强制性简化是指由两种语言系统差异所造成的显化；非强制性简化是指由翻译过程本身、译者偏好、目标语语文规范等因素造成的显化(Baker 2004:82-83)。

第四，对各种变量加以关注和控制。翻译共性研究结果不尽统一的一个原因就是各种变量的存在。其中，语料库设计中的可比性缺乏问题、研究中的参数选择与文体类型选择、翻译方向、译者因素等会对研究结果产生很大影响。

第五，从语言、认知、社会文化等视角对各类翻译普遍现象进行解释。Chesterman(2004b:10-11)指出共性若的确存在，就应当有其存在的原因和影响，因此在对翻译共性进行充分描写的基础上，不仅从翻译活动内部，如语言，更应从外部，如认知和社会文化等视角对其做出合理的解释。柯飞(2005)在考察汉英/英汉翻译中的隐和显时指出：翻译中的隐和显可由语言、译者、社会文化、文本类型等多种因素造成。若将这些方面综合起来进行考虑就可以尝试建立一套对翻译普遍现象的解释机制，加深对于翻译本质的认识。

第六，对翻译共性研究成果的应用。既然翻译语言有其特殊的存在形式，我们在实践中应当遵循还是抵制此类特征呢？从逻辑上讲，翻译语言的普遍特征是翻译文本有别于原创文本的主要所在，如果将原创文本作为常态的语言的标志，那么翻译文本中语言便是对常态偏离，因而常被冠以"翻译腔"、"杂合语言"、"第三类语码"等名。实际上，各种语言变体之间没有孰优孰劣之分，所谓的褒贬都是在一定社会文化条件下的人为因素造成的。翻译语言作为一种客观存在的"中介语"，其特征不应被忽略，也不应被夸大。一定时期内翻译语言普遍特征的累积是这一时期翻译概念和翻译规范建立的基础，认清此类特征有助于我们在翻译研究中更好地认识特定时期的翻译规范。

Halverson(2003)认为翻译共性研究所涉及的各种词汇和语义模式都可参照人类认知的一般特征加以解释。她借鉴 Croft(1990)对语言共性研究的归纳方法，将基于语料库的翻译共性研究也划分为三个层次：1)描写层面，以平行语料库为基础，首先对具体语言对中与翻译相联系的各种关系和结构加以描写；2)(内部)归纳(internal generalization)层面，对各类现象从语言内部及跨语言的角度进行归纳，单语类比语料库的研究即属于这一层次；3)解释层面，即外部归纳(external generalization)，从人类认知心理学、社会学和文化的角度对各类翻译共性做出更高层次的解释。Halverson(2003)认为，目前的翻译共性研究尚处于第二个层面上。

第7章

英译汉语言特征宏观描述

本章基于汉英双向对应语料库在宏观层面描写和分析英译汉语言的总体特征。语料库考察表明：与汉语原创文本比，英译汉文本语言的存在类符/型符比较高，句段偏长；汉语翻译语言扩大了汉语某些结构式的容量等特征。这些特征呈现出英译汉文本独有的特征，显示英语源语对汉语译文的影响，同时也在一定程度上对翻译共性的假说提出了挑战。

7.1 引言

随着计算机语料库技术的发展，语料库语言学已经从方法论和语言研究手段成长为一门完整的学科，有纯理论研究也有应用研究(Laviosa 2002)。基于语料库的翻译研究，可称为语料库翻译学，其主要含义就是，以语言理论和翻译理论为研究上的指导，以概率和统计为手段，以大规模双语真实语料为对象，采用语内对比与语际对比相结合的方法，对翻译现象进行历时或共时的描写和解释，探索翻译的本质(王克非、黄立波 2007)。总之，基于语料库的翻译研究已成为一个独特的研究领域，可以从更多的角度透视翻译，包括翻译语言。

从一种语言翻译到另一种语言多少会存留一些源语的痕迹，有人称之为"translationese"，即受源语影响的不太地道的目的语运用方式(Hopkinson，2007)。"Translationese"若译为"翻译腔"则带有些贬义，译为"翻译体"稍好，也可以使用"translational language"这个相对中性的术语。翻译语言虽然不够"地道"，但它通常不是语法错误(mistake)，多属语用上的"偏误"或差误(error)，源语是这些"偏误"的源头(Toury 1995：208)。用翻译语料来研究翻译语言(Gellerstam 1986：88-95)已经走过20多个年头。

根据Santos(1995：60)，翻译语言研究包括三个方面：一是所有翻译产品的属性，即翻译共性(universals of translation，Baker 1993)；二是特定语言间翻译产品的属性，即翻译语言(translationese)的属性；三是某个或某类翻译产品反映出来的语言特征。本章主要考察英语翻译成汉语后汉译语言的特征，这方面的研究对于

英汉语言对比也有重要意义。

7.2 基于语料库的翻译语言特征研究

翻译语言现象早有人注意。Gellerstam(1986)对比瑞典语翻译和非翻译小说文本,从语内类比角度讨论过"翻译体"(translationese)的语言特征。近年在各种语料库的基础上,开展了一系列新型翻译研究。国际上,英国曼彻斯特大学科技学院的 Mona Baker 较早开展基于语料库的研究。她在"语料库语言学和翻译研究:启示与应用"(1993)一文中对这两者的结合作了初步阐发。在大范围翻译调查方面,如翻译文体的考察,她(2000)也率先从语料库角度探讨译者的文体特征,特别是从类符-形符比、平均句长及词项使用特点等方面加以分析。根据 Baker(1993, 1998a),翻译语言可能存在下列共性:简化(词汇上的简化,即使用较少的词来进行翻译,尽量少用生僻的字眼,使用释义方式应对文化差异;句法上的简化,即复杂结构被简化;文体简化,即译文倾向于断开长句,减少或省略重复的冗余信息,减少词汇重复等)、显化(原文隐含的内容在译文中被明示出来,衔接手段显化)、规范化(译文尽量向目的语的某些典型样态靠拢,缺少目的语使用的多样性)。

Baker(1998b:225)认为,翻译时与源语文本特征相关的现象倾向于移植到目标语文本里。但有批评者指出,Baker 的研究主要基于类比语料(comparable corpus),未关注源语对译语的影响(Hansen & Teich 2001;吴昂、黄立波 2006;黄立波、王克非 2006)。

基于对应语料库的汉语翻译语言研究在国内也已取得不少进展,如柯飞(2003);秦洪武、王克非(2004);柯飞(2005);王克非、胡显耀(2008);等等。王克非、胡显耀(2008)依据双语对应语料库仔细探讨了翻译汉语的特征,指出:无论是文学还是非文学语料,翻译汉语与原创汉语相比,具有词语变化度偏低、词汇密度偏低、虚词显化、指代方式显化、常用词频率增加等特征。国内外这些基于语料库的研究都不同于传统的"对等"理念指导下的翻译研究,而关注起翻译转换中语言之间的相互作用带来的语言变化以及促成此类变化的语言内、外因素,开辟了新的翻译研究途径。但囿于检索技术和研究手段的局限,系统的描述和分析还不够。本章尝试运用现有的语料检索和分析技术,从宏观的层面(主要是类符/型符比、句长分布、结构容量等)探讨英译汉的一些语言特征,微观语言描写(词语分布、词语组合等)部分见第8章。

考察汉语翻译语言,实际上就是考察它在多大程度上偏离汉语"常态"。"常态"(余光中 2002:151)只是一种语言直觉,并无具体标准,但我们可以假定它存在。比如可以把一定规模的原创文本视为汉语"常态",因为它比翻译文本更接近

目的语的基本特征。

本研究的素材取自北京外国语大学研制的汉英对应语料库(CEPC)[①]，我们从中抽取约350万字/词的双语对应语料。该库为汉英双向翻译对应语料库，各子库里的汉语文本可以提取出来，形成汉语原创文本和汉语翻译文本类比库。既然翻译是由源语语篇诱导的篇章产生过程，在进行具体语言分析时，我们既考察汉语原创文本和汉语翻译文本之间的关系，又考察汉语翻译文本和英语源语文本之间的关系。同时，我们也注意文学文本与非文学文本的语言使用有多少差异。

7.3 对英译汉语言特征的宏观描述

宏观描述主要涉及汉语原创文本和翻译文本在类符/型符比、词长和平均句长上的表现。类符/型符比是类符(types)和型符(tokens)之间的比率。词型和实例依据ICTCLAS进行切分，也是句长的计算单位。表7.1是使用WordSmith 4.0得到的统计结果：

表7.1 原创文本和翻译文本的宏观描述

	统计型符数	类符数	STTR	词长	平均句长	平均句段长
汉语原创文学	466 414	23 047	46.72	1.36	25.46	6.02
汉语原创非文学	222 758	11 066	41.92	1.76	27.05	7.20
汉语原创文本	689 172	28 437	45.19	1.49	25.95	6.35
汉语翻译文学	578 148	24 213	47.36	1.44	25.81	7.00
汉语翻译非文学	496 218	26 174	47.65	1.64	31.52	8.58
汉语翻译文本	1 074 366	36 354	47.49	1.53	28.27	7.65
英语原创文学	546 632	22 409	43.21	4.26	16.76	6.79
英语原创非文学	487 673	25 739	44.37	4.87	20.24	9.32
英语原创文本	1 034 305	35 695	43.75	4.54	18.23	7.78

表7.1反映的英译汉语言基本特征可分析如下：

7.3.1 类符/型符比

一般说来，语库规模越大，类符(type)型符(token)比越小。若语料库规模相

① CEPC(Chinese-English Parallel Corpus)是一个汉英双向平行对应语料库，于2002～2003年由王克非教授主持创建，语料以XML文件格式存储，收中英文近3 000万字/词。

同，类符/型符比越大，就说明用词越丰富。由于原创库和翻译库库容差异较大（如汉语原创文本库在实例数量上小于汉语翻译文本，约为 1:1.6），简单的类符/型符比计算没有说服力。因此，本章在进行类符/型符比（type/token ratio）计算时，使用了 WordSmith 4.0 提供的以 1 000 词为计算单位的类符/型符比计算方式，得到"标准化类符/型符比"（见表 7.1）。

表 7.1 显示，在类符/型符比上，汉语原创文本比汉语翻译文本在类符/型符比上低了 2.3 个百分点（45.19 vs. 47.49），但都高于英语原创文本（43.75）。汉语翻译语言在用词丰富程度上低于英语源语，但又高于汉语原创语言。这一差异说明，汉语翻译文本在用词丰富程度上高于汉语原创文本和英语源语文本。该项统计结果不支持翻译语言较原创文本用词简单这一假设。

7.3.2　平均词长

总体说来，汉语原创文本和翻译文本都以双音节和单音节词占主导地位，汉语翻译语言基本符合汉语本身的词汇运用特征。汉语原创文本使用单音节词稍多，但总体看来，两者没有明显的区别。

7.3.3　平均句长

统计显示，汉语翻译文本的平均句长高于汉语原创文本（每句平均高出 2.32 个词）。这一差异在非文学文本里差异尤为明显（高出 4.47 词），差异不大的是文学文本（差值：0.35）。那么，这些差异和英语语源语文本有什么关系？仅从句长上说，汉语翻译文本句子偏长和英语源语似乎没有多大关系，因为后者的句长（18.23）远远小于前者。但实际翻译的语言材料显示，汉语翻译时的句子扩张与英语源语有关。

(1) And Pinkerton—Pinkerton—he has collected ten cents that he thought he was going to lose.

那么/c 平克顿/nr—/x 平克顿/nr—/x 他/r 一定/d 是/u 要/v 回来/dg 一/m 角/q 钱/n 的/u 老/a 账/n，/w 这笔/r 钱/n 他/r 本来/d 以为/v 没有/v 盼头/n 了/y。/w(16:22)

(2) You speak collectedly, and you—are collected.

你/r 这/r 话/n 倒/y 还有/v 自制力/n，/w 而/c 你/r—/x 也/y 确实/d 镇静/a。/w(7:10)

汉语的孤立语性质意味着要使用词汇手段表达一些英语使用语法或其他组合

手段表达的内容。如例(1)中的关系词 *that* 译为名词短语“这笔 钱”，过去式(thought)在翻译时通过词汇手段(本来)添加时间标记(译为“本来 以为”)。还有，英语介词(如 on、in)表达的意念在汉语中常用“介词+……+方位词”来表达。上述各种添加多少带有一些强制性，更常见的添加则是选择性的。比如，翻译时可能会发生词性改变(如例(2)中的“这、话”、“自制力”)，相应的成分都须作出调整，句子调整的幅度随之加大，原文中隐含的表情态成分也可能随之明示出来(如“倒、也、确实”)。不论是哪类添加，都会致使汉语翻译语言的平均句长高于英语原创文本语言。这可以部分证实显化现象。

但这只反映出汉语翻译语言的句子长度，而仅计算以句号、问号和叹号为切分点的句子长度还无法反映句子内部组织特征。根据 Chen(1994)，75%的汉语句子由两个以上的句段构成。句段可以是一个小句，也可以是一个短语，但它一定是句中由逗号或分号隔开的相对独立的构句成分。句段统计显示，汉语文本的句段长度低于英语文本；更重要的是，汉语翻译文本的句段长度明显高于汉语原创文本，且更接近英语源语文本。

仅从统计数据上看，汉语翻译语言用词比较丰富，句子偏长，没有期待中那么通俗易懂，也没有表现出翻译共性所言的句法简化特征。而且，汉语翻译文本中的句段长度大于汉语原创文本，这显然是英语源语干预的结果。

7.4 结构容量

根据表 7.1，汉语翻译语言使用的句子偏长，原因很多。从语法上说，由于汉语是不同于印欧语言的非形态语言，英语里的一些语法手段(如英语里的体和时态)在汉译时通常需要运用词汇手段来表达，这自然会增大句长；从组织方式上看，英语常使用较长的分句，在汉译时若保留分句形式也会增加句长；从翻译策略上看，汉语译文可能会为了方便目的语读者理解而加上一些解释性话语。

但语料分析显示，上述差异并不足以说明汉语的翻译和原创文本在句长上存在明显差异，翻译语言扩展了汉语某些结构式的容量才是主因。容量扩展表现在以下两个方面：

谓语和宾语之间的距离拉大了，即英汉翻译时常把原本后置且开放的英语定语(从句)前置于宾语之前(主要是“的……的 NP”结构式)，这大幅提高了汉语谓宾之间的结构张力；

汉语的一些介词短语(介词-NP-方位/时间)本身是封闭结构，容量有限，但翻译语言通常需要将较长的多重修饰成分塞入这种结构，致使其容量扩张。

我们以“在……中”为例说明这一点(见表 7.2)。

表 7.2　“在……中”的结构容量

	原创非文学		原创文学		翻译非文学		翻译文学	
	频次	相对频率%	频次	相对频率%	频次	相对频率%	频次	相对频率%
4 个词以内	260	67.88	183	91.5	941	72.33	448	74.91
5 个词以上	123	32.12	17	8.5	360	27.67	150	25.09
合　计	383	100	200	100	1 301	100	598	100

如表 7.2 所示，“在……中”结构的容量与文本类型相关。在汉语原创文学文本中，有 91.5%的“在……中”结构的容量在 4 个词以内，单音节词更是占到了 47%；而在翻译文学文本里，结构容量在 4 词以内的只占 74.91%。可以看出，原创文学文本里“在……中”结构的容量要远远小于原创非文学文本。有趣的是，这一区别在两类翻译文本中表现得不是那么明显，原因可能是汉语文学翻译不太注意该类结构式在语域上的差异。

与“在……中”结构相似的还有“在……时”和“介词＋NP＋处所词”结构。这些封闭的结构式容量有限，但在翻译语言里，它们的容量均有大幅扩张。

上述扩容现象受源语影响，但不是源语单独作用的结果。还有一个原因：翻译语言过分利用了汉语本身的结构张力。尽管英语谓-宾之间容量有限，但修饰成分后置（如使用定语从句、介词短语）可以避开这种容量限制。汉语则不然，它的修饰成分通常前置。所以，若不使用其他方式，汉语只能靠扩展结构容量来解决修饰问题。扩容意味着不能快速处理主干成分，这会加重工作记忆负担。因此，结构容量扩张后，汉语翻译文本语言不如汉语原创文本语言易读、易解。

7.5　小结

汉英对应语料库的考察和分析表明，同汉语原创文本比，英译汉文本语言的标准化类符/型符比较高、句段偏长，表明汉语翻译文本在用词丰富程度上高于汉语原创文本和英语源语文本。另外，汉语翻译语言扩大了汉语某些结构式的容量，导致汉语翻译文本语言不如汉语原创文本语言易读、易解。这些特征呈现出英译汉文本语言独有的特征，显示英语源语对汉语译文的影响。从这个意义上说，特定语言间的语言转换现象，或翻译语言的局部特征，不一定是所有翻译语言共有的属性，翻译共性的探讨还有待深入。

第8章

英译汉语言特征微观描述

本章基于汉英双向对应语料库从微观角度描写和分析英译汉语言的词汇特征。研究发现：与汉语原创文本相比，汉语翻译文本在词类分布和某些词(素)的组合能力上有差异；在词类分布方面，汉语翻译文本较汉语原创文本更多地使用虚词，较少使用实词；汉语翻译文本比汉语原创文本较少使用单音节词、较多使用双/多音节词，其介词、连词和代词的使用频率也高于汉语原创文本；在词汇组合方面，英译汉文本可能使用汉语原创文本不太常用的词或多词组合形式。这些特征大多为英译汉语言独有的特征，不能完全用翻译共性来解释，而应充分考虑英文源语对译文的影响。

8.1 引言

本章研究的是特定语言间翻译产品的特性，即翻译语言(translationese)的特征，即 Santos(1995:60)关于翻译语言研究讨论的第二个方面，具体就是英译汉语言的特征，这类研究对汉英语言对比有重要意义。前一章在比较宏观的层面上探讨了英译汉语言特征，本章侧重翻译语言的微观层面，即尝试运用现有的语料检索和分析技术在词类分布和词语组合等能够体现语言个性的层面上描写和分析英译汉语言的基本特征。

本研究使用的素材和例句取自北京外国语大学的通用汉英对应语料库(CEPC)。该库为汉英双向翻译对应语料库，各子库里的汉语文本可以提取出来，形成汉语原创文本和汉语翻译文本类比库。由于翻译是源语语篇诱导的篇章产生过程，因此，在进行语言分析时，既对比汉语原创和汉语翻译文本，又考察汉语翻译文本和英文源语文本之间的关系。

8.2 英译汉语言的词类分布

8.2.1 统计数据

词类分布能部分反映语言的类型和特点。表 8.1 是各主要词类在 CEPC 中英

文原创和翻译文学子库中的分布情况。汉语分词和词类标注使用 ICTCLAS(北大标准),英文词类标注使用 CLAWS。

表 8.1　英译汉词类分布频率

词类＼分布	汉语原创文学		汉语翻译文学		英文原创文学	
	词数	频率	词数	频率	词数	频率
动词	110 391	23.64	133 762	22.93	108 340	19.88
名词	100 827	21.59	113 823	19.52	112 536	20.65
形容词	24 948	5.34	24 672	4.23	35 846	6.58
副词	48 676	10.42	52 266	8.96	42 065	7.72
代词	41 259	8.83	68 859	11.81	64 433	11.82
介词	14 536	3.11	25 932	4.45	58 743	10.78
连词	9 687	2.07	15 252	2.61	39 304	7.21
数词	17 322	3.71	20 174	3.45	8 463	1.5
量词	14 209	3.04	16 337	2.80	0	0
助词	39 370	8.51	57 372	9.84	0	0
冠词(英文)	0	0	0	0	52 325	9.60
限定成分(英文)	0	0	0	0	17 132	3.14
总词数及所占比率	467 007	90.26	583 260	90.6	545 000	98.88
备注	其他词类未作统计		其他词类未作统计		其他词类未作统计	

根据表 8.1 制作的词类分布折线图见图 8.1。

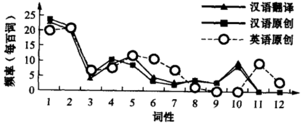


图 8.1　词类分布折线图

注:横轴上的数字分别代表:1=动词;2=名词;3=形容词;4=副词;5=代词;6=介词;7=连词;8=数词;9=量词;10=助词;11=冠词(英文);12=限定成分(英文)

从图 8.1 中可以清楚地看出,汉语翻译文本和汉语原创文本的词类分布线型

基本一致,但和英文原创文本差异较大。可以说,汉语翻译语言与汉语原创语言之间的差异要远远小于它和英语源文之间的差异,即汉语翻译语言总体上符合汉语的语言特征。

从图中还可看到英文动词的使用频率要低于汉语近 4 个百分点,这可以证实英文多用名词而汉语动词活跃的观点(思果,2002:55-58;邵志洪,2005:24)。但英文代词、介词和连词的使用频率则高于汉语(分别高出 3、7、5 个百分点)。受这一差异影响,汉语翻译语言在动词使用上略低于汉语原创文本使用的语言,在代词和介词使用上则高于后者。但在连词使用上,汉语翻译语言似乎没有受英文太多的影响。

汉语翻译文本和汉语原创文本在词类分布上有以下特征:前者动词、名词、形容词和副词的使用频率低于后者;而在代词、介词、连词、助词等的使用频率上明显高于后者。换言之,汉语翻译文本使用实义词的频率较低,使用功能词的频率偏高。鉴于英文原创语言中功能词的使用频率明显高于汉语原创语言,有理由认为这一差异是翻译时受源语即英文影响的结果。

目前的分析说明,汉语翻译文本和汉语原创文本在某些词类的使用频率上有差异,但词类分布的形态基本一致。然而,仅凭这些还不足以充分反映两类文本在词汇使用上的差异,我们还需要考察词类内部某些词项使用的情况。为此,本章使用 WordSmith4.0 中的 Keywords 工具对比翻译文本和原创文本在词汇使用上的差异。以下是对比结果。

8.2.2 实词的使用

8.2.2.1 名词

两类文本在名词使用上的差异表现在以下两个方面。词汇方面:汉语原创文本高频使用当差、洋车、饺子、表姐、姑奶奶、旗袍、夜壶等词语;而汉语翻译文本高频使用上帝、绅士、牧师等词语。名词素方面:汉语原创文本高频使用兄、时、寺、堂、氏、帖、斋,涉及称谓(如令堂)、时间(如午时)、命名(如白塔寺)。翻译文本和非翻译文本在普通名词使用上的差异大都与社会、文化和生活环境有关,有时凭直觉就能判断出来。

8.2.2.2 动词

统计显示,汉语翻译文本中有几类词的使用明显高于原创文本:1)体动词:开始、结束;2)发生类动词:发生、产生;3)发现类动词:表现、发现;4)使让类:让;5)判断类动词:认为、相信、感觉;6)情感类动词:害怕、怀疑;7)其他:具有、存在。这些词在汉语原创文本中的出现频率也很高,但远没有翻译文本用得频繁。原因只有一个,与这些词对应的英文表达(动词、介词短语、形容词)在英文源语文本里均高

频使用。

统计还显示，在汉语原创文本里高频使用的单音节动词远远多于汉语翻译文本，如单音节动词凑、搁、甭、傍、嚷、嫌、吵、捧、混、怔。该现象可以这样解释：论数量，汉语里双音节词比单音词多，但最常用的词多是单音节（吕叔湘 1999[1980]：10），而从外文翻译成汉语的词通常都是双音节词。

8.2.2.3　形容词

对比表明，原创文本和翻译文本在形容词使用上的最大差异是单音节形容词的使用。例如，原创文本中高频使用脆、高、贵、好、红、厚、慌、紧、老、俏、小、饱、苦等单音节形容词，而翻译文本中高频使用的单音节形容词要少得多，只有大、久、多、太、快等。很明显，类似于单音节动词的使用，单音节形容词的使用数量和使用频率也能说明汉语翻译文本使用双音节词的倾向更强一些。

8.2.2.4　代词

在代词使用上，原创文本使用的下列代词明显高于翻译文本，如：大家、她们、怎、怎样、这、自己等。而翻译文本高频使用的代词和指示词有：她、他、他们、它、它们、我、我们、那、那儿、那个、那时、那种、这个、这时、这种、这样、其他、另、别的、任何、每个、一切等。不难看出，相对于汉语原创文本，翻译文本高频使用第一和第三人称代词、指示代词和量词，这是汉语翻译语言的独特之处，有关的分析见 8.3.2 节。

8.2.2.5　方位词

汉语原创文本高频使用的方位词不多，只有里、外边、内。而汉语翻译文本高频使用的方位词十分丰富，如：以前、之前、之间、之后；之中、之下、下面、附近、上面、之外、中、周围。出现这种差异的主要原因是源语英文里高频使用介（连）词，如 before、after、between、in、under、near、around 等，翻译文本受此影响而使用较多的双音节方位词。

8.2.2.6　副词

汉语翻译文本在副词使用上有以下特点：

翻译文本高频使用的时间副词远远多于原创文本，主要是因为英文原文里的语法手段如完成体和进行体在译成汉语时需要以词汇手段（正、已、已经、一直）表达出来。

语气类副词表达说话者对某一命题的态度，翻译语言中使用频率相对较高的副词有必须、或许、竟然、大约、如此等。原创文本也高频使用一些表达情态的副词，如得、兴许、原来、却、来（如十来天）、这么等。因此，这种差异不只是频率上的，还体现在遣词方面，即汉语翻译文本在副词使用上变化较少且有双音节倾向，不如原创文本那么丰富。

8.2.3 虚词的使用

在虚词方面，汉语翻译文本和汉语原创文本的差异主要表现在连词的使用上。统计显示，汉语翻译文本高频使用的连词在总量上多于原创文本(16∶7)，有句间连词(不过、但、但是、尽管、或者、不仅、而且、另外、哪怕、即使、如果、然后、因此、于是)和词语间连词(和)。与之相比，原创文本中高频使用的连词为数不多，只有“可是、并且、况且、不但、所以、假若、愈……愈”。出现这一差异的原因是：汉语常隐含逻辑关系，而英文则更倾向于将逻辑关系明示。正是受到源语英文的影响，翻译文本较原创文本更倾向于使用连词来表达各种逻辑关系。

上面的分析显示，汉语翻译语言在词类分布上深受源语的影响，表现在两个方面：一、汉语翻译文本较汉语原创文本使用功能词偏多，使用实义词偏少；二、汉语翻译语言倾向于使用双音节词，单音节词使用的数量和频率都不及汉语原创文本。

8.3 英译汉词语的组合方式

汉语翻译语言可能会过度运用目的语某些语素或词汇的组合能力，导致这些词(素)的组合能力变强，表现在以下三个方面。

8.3.1 名语素的使用

下面以“～性”为例对比汉语名语素的组合能力。从分布密度上看，汉语原创文学文本中“～性”类名词的分布频率约为 2.9/万词；而它在翻译文学文本中的分布频率为 5.2/万词。不仅如此，“～性”在翻译文本中组合方式多样。例如，翻译文本中使用了 71 个“～性”组合词，其中有 42 个是原创文本里低频使用的：如独创性、决定性、历史性、可信性、坚定性、实质性、强制性、煽动性，等等。与翻译文本相比，原创文本中“～性”组合词的数量较少(共 41 个词)，且多与单音节词组合，如爽性、土性、火性、牛性、癖性、韧性，等等。

这种组合上的差异显然与英文原文-lity，-ness，-dom，-ity 类后缀的高频使用有关。在处理这类转类词时，汉语翻译语言倾向于使用“性”语素附加类别标记，最终导致其组合能力增强。

上述情形适用于“～力”、“～度”等组合。这些带有术语特征的名语素在翻译文本中的出现频率上高于汉语原创文本，很大程度上是对源语组合方式的模仿和借用。

8.3.2 “指示代词+数量词”组合

检索显示，翻译文本高频使用指示代词(这/那)和数量短语组合。表 8.2 显示

的是差异最明显的三个组合。

表8.2 指示代词+数量词组合在两类文本中的分布

	翻译文本		原创文本	
	频 次	转移几率①	频 次	转移几率
这一	1010	0.035	334	0.019
这种	491	0.09	39	0.012
这件事	123	0.017	81	0.018

"这一":原创文本中与"这一"有统计意义的搭配词有点(71)、条(19)、天(11)、次(28)、年(10);翻译文本里除了条、天、次,还有"问题、条款、事实、目标、领域、类、计划、过程、观点"。差异很明显:在原创文本里,"这一"与单音节量词搭配居多;在翻译文本里,它与双音节普通名词共现得更频繁。

"这种":"这种"在原创文本里的组合能力不强,主要和"人"共现;即便如此,为数也远远少于翻译文本(分别为39和295)。在翻译文本里,它的组合方式异常丰富,还频繁跟药、病、事、做法、想法、现象、感觉、情况、方式、方法等名词组合在一起。统计显示,翻译文本中"这种、这一"的高频使用与英文源语 *the/this/that* 的高频使用密切相关。

出于同样的原因,翻译语言中数量短语(一个、一件、一位、一片等)的使用频率远远高于原创文本,主要原因是:汉语译文倾向于以显性的表达(数量短语)来处理冠词这个英文中高频使用的语法范畴。

当然,有些数量短语的使用还与它在语篇中所起的衔接功能有关。比如,"这件事"在翻译文本和原创文本中的转移几率相当。它在翻译文本中的出现频率虽高,但在英语源语中与它对应的大都不是 *the matter*, *the thing*,而是 *it*, *this*, *that*。如:

(1) a. It had happened at last.
　　这件事终于发生了。
　b. "Nobody knows about this but us?"
　　"除了咱们,没人知道这件事吧?"
　c. "Humph! We'll see about that."

---

① 转移几率(transitional probability)衡量某一词穿在给定文本中一起出现的概率。例如"这"在给定语料库中出现100次,其中2次出现在"一"之前,而98次出现在其他词之前,那么,转移几率 p(一|这)=2/100=.02;而转移几率 p(其他|这)=98/100=.98。转移几率的计算方式类似于MI值,值越高在一起出现的几率就越大。该数值来自 Antconc 检索软件。

"嗯，这件事我们得管一管了。

d. There must "a" been an angel there.

这件事一定有个高手在帮你的忙。

从语篇角度看，上述例证改变了衔接方式(即名词短语替代了源语中的代词和有指示意义的副词)。改换衔接方式显然反映了中英文在衔接手段上的差异(汉语较英文更多使用词汇手段)，但结果导致了"这件事"在汉语翻译语言中的高频使用，所以导致这一状况出现的原因还是源语。从另一个角度看，转换衔接方式或添加衔接成分都是为了让译文更符合目的语规范。这种显化现象在翻译语言中有，但不多；总体看来，汉语翻译语言代词使用的密度要高于汉语原创语言，接近英文源语。

以上分析说明，源语的一些语法特征(冠词、指示代词)驱使译者不自觉地模仿英文源语的表述方式，向源语靠拢。其结果是，汉语原创语言中的某些组合方式被过度使用。

8.3.3 表达式用法相对固化

汉语翻译语言常常会扩展汉语中某些表达式的功能和用法，或者高频使用某些目的语里的结构形式。根据柯飞(2003)的研究，汉语翻译文本中使用"把"字句的频率要远远高于汉语原创文本，主要原因如下：1)英文里高频使用一些表达客体位置变化的动词，如 *put*，*place*，*push*，*hand* 等，这些动词的"处置"义正好可以使用"把"字句来表达；2)英文里有些常用动词，如 *turn*，*break*，*smash*，*clean* 等，融合过程与结果，通常表达宾语受到外力影响发生位置或状态改变，也适合用"把"字句来表达；3)英文里有些表处置义的结构式如"*have... done*，*get... done*，*let/make...*"也可以译成"把"字句。上述语词在英文源语中高频应用是汉语译文里高频使用"把"字句的主要原因。

Schiffrin 认为，英文里的话语标记如 *I think*、*it seems* 等评述性小句(comment clause)和日常交际用语(conversational routine)如 *thank you* 等的词汇化程度较高，原因是它们的固化程度较高(转引自 Brinton & Traugott 2005:67)。在汉语翻译语言里，受源语影响的某一典型译法在翻译时变成了相对稳定的表达形式，用于表达源语中其他有相近意义的表达形式，这就形成了表达式的固化。限于篇幅，本章只介绍两个。

"随着时间的推移"

该表达式在翻译文本里可谓“一专多能”。一专是指它专用于表达时间流转；多能是指它能对应源语中多样的时间表达，如 as time went on, moment by moment, over time, as time drifted along, with a long-term time horizon, in the course of time, as time went by，甚至还有 eventually。

另外，这个表达式已经成为汉语里的常用表达式。虽然在汉语原创文本里仅检索到两个（随着季节的推移、随着时间的推移），频次不高。但在 CCL 语料库中，却检索到 503 次（其中在非翻译作品中出现 402 次）。这说明，随着时间的推移已经成为汉语表达时间变化的常用形式，它的频繁使用应归功于翻译。

“是（不）可能（的）”

“是（不）可能（的）”的主要功能是表达命题态度，它在汉语翻译语言中的使用频率明显高于原创文本。表 8.3 是英汉翻译子库中的检索结果。

表 8.3 “是（不）可能（的）”结构在英文源语中的对应项

原语对应词	POSSIBLE	CAN	LIKELY	WILL	Might	Probable	Incapable	无对应	合计
频次	41	10	5	2	2	1	1	3	67

在原创文本库里，“是（不）可能（的）”只出现 17 次，均出现在政论文本里。而且，该结构式多以“是（不）可能的”形式置于句末，很少使用扩展式。总之，这个结构在汉语中使用频率低，适用范围窄，使用方式也比较单一。

在翻译文本里，“是（不）可能（的）”对应于英文 POSSIBLE 类语词。这种对应性可以扩展到其他与之意义相近的词项上去，用法和适用范围也随之变大（可以对应 *can, will, might, incapable*）。另外，“是（不）可能（的）”在句中使用扩展式的频率远远高于它在原创文本中的使用频率。比如，在下面的汉语译文里，该结构被大幅度扩展，这一扩展显然是受到了源语句子铺排方式的影响：

(2) Earlier in life, Hester had vainly imagined that she herself might be the destined prophetess, but had long since recognized **the impossibility** that any mission of divine and mysterious truth should be confided to a woman stained with sin, bowed down with shame, or even burdened with a life-long sorrow.

海丝特年轻时也曾虚妄地幻想过，她本人或许就是命定的女先知，但从那以后，她早已承认了：任何上界的神秘真理的使命是不可能委托给一个为罪孽所玷污、为耻辱所压倒或者甚至为终生的忧愁而沉闷的女人的。

汉语翻译库中还检索到另一些高频使用的表达式，如目的是为了、在某种程度/意义上、……是必要的、很久很久以前、更确切地说、一般情况下。这些表达式在汉语原创文本里也有使用，但使用的频率远远低于汉语翻译文本。

不难看出，在翻译文本中一些高频运用的表达形式，一旦被广泛接受和认可，就可能进入目的语，成为多个近义表达式的翻译对应项。这说明，翻译语言倾向于使用通用的表达方式处理源语中形式不同但功能相似的表述，这可以部分证实翻译共性中的词汇简化现象。

8.4 小结

语料考察和分析表明，汉语翻译语言具有以下特点：1)汉语翻译语言高频使用的单音节词较原创文本少；2)汉语翻译语言里的介词、连词和代词在使用的数量和频率上高于汉语源语文本；3)翻译语言完全可能使用原创文本不太常用的词或多词组合形式，因此不能简单地认为原创文本在词汇运用上不如目的语文本丰富。

本章的统计和分析结果对翻译共性说构成质疑。首先，汉语翻译语言局部的显化与一些词项(如关系代词、逻辑连接词、代词等)的隐化并行。而且，显化是相对的，就英汉翻译而言，汉语翻译语言相对于汉语原创文本而言显化，相对于英文源语而言隐化。显化的方式和程度显然与源语和译语本身的性质有关，它是相对的，同时又具有局部特征。其次，汉语翻译语言的整体特征虽然与汉语原创语言一致，但前者结构容量扩增，句子偏长(见第7章)，因此不完全支持规范化。再次，翻译语言用词也很丰富，特别是高频使用源自源语的一些表达方式，而且句段较原创语言长，也体现不出翻译共性里的简化特征。

第9章

翻译特征分析:显化与隐化

近年的翻译研究开始探讨翻译之共性,翻译中的隐化和显化这类特征可能涉及翻译共性。本章以英汉互译实例分析翻译中显化和隐化现象的发生,认为显、隐现象由语言、译者、社会文化等多种因素造成,并着重讨论了语言因素和译者因素。分析发现隐、显现象的发生和隐、显程度可能与语言的形式化程度和翻译方向相关。

9.1 翻译的特征与共性

译文是否忠实于原文,这可以说是检验翻译优劣的永恒标准,也是翻译研究、翻译评论的一个主要话题。但语料库特别是双语语料库的出现,引发出人们对翻译研究的新思考。例如近十多年里国外学者致力探讨的翻译之共性问题,以及与之相关的译文语言特点问题。这是翻译研究中的一个新视角。语料库的发展为这一类翻译研究提供了更充分的资源和手段。

翻译的共性,即译文中呈现的有别于原文的一些典型的、跨语言的、有一定普遍性的特征。近年来涉及语料库与翻译的研究报告,有不少是围绕这一课题展开的,如以下值得注意的文献:

(1) Baker(1993,1998)通过对"英语翻译语料库"(TEC)的考察,提出译文语言呈简化(simplification)特点;

(2) Laviosa(1997,1998)考察译文与母语在词汇使用上的不同,发现译文比母语使用更多的高频词;这似乎可佐证 Baker 的观点,因高频词数量少,使用的多意味着词汇简化;

(3) Øverås(1998)通过对英语和挪威语互译的考察,证明译文语言同母语相比,有显化(explicitation)倾向;

(4) 王克非(2003)基于对汉英对应语料库的考察,发现译本扩增现象,这与Øverås(1998)的考察结果类似;

(5) Kenny(2001)基于德语/英语文学素材平行语料库的考察,发现创造性新

词翻译中呈现标准化/规范化(normalization)趋势;

(6) Kenny(1998)通过对原文、译文的语义韵分析,发现同母语相比,译文语言有"净化"(sanitisation)现象;

(7) 但是译文也许不仅仅有"简化"、"显化"和"净化"等现象,我们认为,(a)翻译常常遇到较曲折的表达,从而可能更多地使用母语中较复杂的结构或句式;通过语料库考察某些结构的分布,我们发现汉语译文比汉语原文使用更多的"把"字句,尤其是文学译本中;(b)翻译过程中会不同程度地发生对原文的模仿,从而使译文变得复杂化、冗长化(柯飞 2003);

(8) 陈瑞清(2004)从自建"大众科学英汉平行语料库"和台湾中研院平衡语料库原创汉语文本对部分汉语关联连词的使用情况作比较,发现汉译本有从意合转形合的潜在显化趋势;

(9) 秦洪武、柯飞(2004)通过个案考察"翻译语"(translationese),亦即既非原文又不完全是母语的、介乎原文和译文之间的"语际语",并注意到非文学译本比文学译本在结构上更趋向于借鉴原文;

(10) 另外,Xiao & McEnery(2002)则通过深入的研究,发现在"体"标记的使用上,汉语译文比汉语原文多出约一倍。

(11) 其他还有 Ebeling(1998)比较英语和挪威语在存现句使用上的特点,Maia(1998)以汉英对应语料库观察英语和葡萄牙语在人称主语使用频率上的差异等(王克非 2004b:182-184)。

对以上涉及翻译共性问题的探讨,无论在国外或国内,还只是初步的,有待深入的,但却展现出广阔的研究前景,有助于我们加深对翻译和翻译过程的认识。

9.2 翻译中的显化现象

显化(explicitness/explicitation),又译外显化、明朗化、明晰化、明示等,指的是"目标文本以更明显的形式表述源文本的信息,是译者在翻译过程中增添解释性短语或添加连接词等来帮助译本的逻辑性和易解性"(Shuttleworth & Cowie 1997:55)。Blum-Kulka 最早对此进行系统的研究,她认为"译者对于原文进行解译的过程可能导致译语文本比源语文本冗长";这一冗长现象"可能是由于译语中提高了衔接上的外显程度造成的"(Blum-Kulka 1986:19)。

这一假说的提出,在西方翻译界引起研究兴趣。为检验显化假说,Øverås(1998)专门做了英语/挪威语双语语料库调查。她的调查涉及 40 个小说片段中的前 50 个句子,其中 20 本是挪威语作品及其英语译本,另 20 本是英语作品及其挪威语译本。她主要研究该假说所涉及的衔接外显程度是否有提高。她同时从英译

挪、挪译英两个方向考察，并排除语言系统内因语法规则引发的显化转移，同时还细致地兼顾到与之相反的隐化现象。

她的发现是，英译挪文本中共有 347 处显化例证，平均每个文本 17.3 处；挪译英文本中有 248 处显化例证，平均每个文本 12.4 处。显见后者少于前者。这些显化更多的是体现在词汇上。如英译挪文本中的 347 处显化，其中 112 处体现在语法上，而词汇方面则有 235 处，占 2/3 强；挪译英文本中的有 82 处是语法上的，词汇上的显化有 166 处，也占 2/3。区别只是显化在英译挪文本中比在相反方向的挪译英文本中更为显著。

这一现象是否为翻译的普遍现象，还需多种语言的进一步论证。贺显斌(2003)就英译汉过程中的显化(显化)现象作了一项实证研究。他选取欧·亨利的短篇小说 *The Last Leaf* 及其汉译本进行显化现象的比较。他考察的结果是：英文小说全篇 134 句，译成汉语后显化程度提高的有 79 句，占 58.96%，也就是说，佐证了 Øverås(1998)，认为在英译汉过程中有显化现象发生。不过，贺的考察还不够充分，其一是仅做了个案考察，篇幅也较小，其二是未考虑译者等因素。尽管如此，英汉语言之间的互译多少存在显化现象这一点应是可以明确的。

9.3　翻译中的显和隐

翻译中的显化现象是我们过去未曾注意的问题，其实也存在与之相反的隐化(implicitness)现象。英语和汉语分属不同语系，形式化程度或形态形式相差很大，因此在不同的翻译方向上，显化或隐化的表现就不一样。我们不妨举例分析。

从英译汉来看，贺显斌(2003)证明了有较高程度的显化，主要是通过使用增词、改用具体词、转换人称、转换辞格等手段。但是鉴于汉语在连接词等衔接手段和指称形式上都不同于英语等印欧语言，所以应该还有一定的隐化现象发生。照顾到汉语表达习惯而做相应隐化处理的翻译通常比不做隐化的仿译(imitation)更为地道。

下面的两例英译汉(本章例句除特别注明外均取自北外的汉英对应语料库)，反映出汉语译者由于受原文影响而产生的仿译，并使译文扩增：

(1) It was one of the few gestures of sentiment he was ever to make.

那是他在感情方面所作出的很少的几次表示中的一个例子。(此译若采取隐化策略，减少重复，避免长定语，则可译得更符合汉语的表达习惯：他很少表露感情，这是难得的一次。)

(2) Can you tell me where is your cereal section?

您能告诉我你们的粮食科在哪里吗？（此译若采取隐化策略，减少不必要的人称代词等，则可译为地道的汉语：请问粮食科在哪儿？）

关于这一点，我们通过回译可以看得更清楚。王克非(2003)曾举贾平凹一篇散文(节选)的英译文及该英译"回译"成汉语的例子。该例显示，因语言结构的需要，译文经显化而扩增，回译成汉语时，篇幅上已比贾平凹原文多出90字，增幅达57%。试分析该片断前头两句：

(3)a 伯父家盖房，想以它垒山墙，但苦于它极不规则，没棱角儿，也没平面儿；(28字)

(3)b *When* Uncle's family was building a house, *we* thought of using it to pile up *a side of the* house wall. *This* proved to be impossible, however, *since* the rock was of an extremely irregular shape, possessing *neither* sharp right angles *nor* any smooth, flat surfaces.

(3)c 伯父家盖房子时，我们曾想到用它垒一面屋墙，但最后还是没能用上，因为这块岩石外形太不规则，既没有尖利的直角，又没有光滑的平面。(56字)

(3)b英译文的斜体部分是因英文表达的需要比较自然地"显化"而添加上去的词语，如连词when、since、neither...nor、主语we等，而且原文隐含的"未能用上该石头"之意也显化出来了：This proved to be impossible。(3)c则顺着英(译)文译成汉语，即(3)b中显化的斜体部分顺译/仿译为(3)c下划线部分；再对照(3)a原文，则可以看出，若回译的(3)c隐化处理(3)b斜体部分，便能基本回到地道的汉语原文。

又如向光(见"翻译工作坊"：http://sts.nthu.edu.tw/transws)批评林宗宪所译《毛泽东》(台北：左岸出版，2001)。原著为美国著名汉学家史景迁(J. Spence)于1999年出版的*Mao Zedong*(New York: Viking Penguin)。向光批评林译为了"省事"，对西文著作中引用的汉语未查找原文，自译引文，结果与毛泽东原文相差甚远。

请看史景迁书中引用的毛泽东《民众的大联合》中一段文字的原文：

(4)a 洞庭闽水，更起高潮。天地为之昭苏，奸邪为之辟易。咳！我们知道了！我们觉醒了！天下者我们的天下。国家者我们的国家。社会者我们的社会。我们不说，谁说？我们不干，谁干？刻不容缓的民众大联合，我们应该积极进行！(收入《毛泽东早期文稿》，湖南出版社，1990，第390页)

(4)b 英译文(史):

From Lake Dongting to the Min River, the tide rides ever higher. Heaven and earth are aroused by it, the wicked are put to flight by it. Ha! We know it! We are awakeded! The world is ours, the state is ours, society is ours. If we do not speak, who will speak? If we do not act, who will act? We must act energetically to carry out the great union of the popular masses, which will not brook a moment's delay!

(4)c 中译文(林):

从洞庭湖到明湖,浪潮更汹涌。天地因而悲愤,邪恶者因而声势高张。啊! 我们知道了! 我们惊醒了! 世界是我们的,国家是我们的,社会是我们的。*假如我们不说话,谁能为我们说话?* *假如我们不行动,谁能为我们行动?* 我们必须有活力地行动以实现广大群众大联合的理想,而那是刻不容缓的事!(林宗宪译,87-88 页)

史景迁不愧为精通英、汉语和中国现代史的专家,其英译文颇显功夫,读来铿锵有力,与毛泽东原文不相伯仲。而林译(4)c 不仅多有错误(如该文中楷体部分),而且"使毛泽东原来力道十足的语言风味尽失"。特别是两个假设条件句(斜体部分),因为有了"假如……"而力道顿失,韵味全无(试比较原文的隐含假设条件句)。

这又令人想到 2005 年 4 月台湾的中国国民党主席连战在北京大学演讲时,以汉语翻译引用美国前总统罗纳德·里根的一段话:"(有些事情)如果我们不去做,谁去做? 如果我们现在不做,什么时候做?"

这句引语带有明显的翻译腔,即对原文的仿译。若采用隐化策略加以改造的话,其译文应是类似上面毛泽东的那段话:"我们不做,谁做? 现在不做,什么时候做?"显化出来的"如果"以及后一个代词"我们"是没有必要的。再看汉译英的例证。

(5)"靠山吃山,靠水吃水,我老汉靠沙子,当然要吃沙!"

"If you live on a mountain, you live off the mountain; if you live by the water, you live off the water. I live on the sand, and I'll live off the sand."

(6)"没有招待所有店,没店有生产队,有老乡窑洞。"

"If there's no guesthouse there'll be an inn; if there's no inn there'll be the cave home of a farmer in the production brigade."

(7) 国家这么大,这么穷,不努力发展生产,日子怎么过?

In a country as big and as poor as ours, if we don't try to increase production, how can we survive?

例(5)至例(7)是汉译英,因英文表达的需要,汉语原文的隐性条件句译成英语时都加上了 if,成为显性条件句。如果按此显性英语条件句回译成汉语(或将类似的英文语句译成汉语),通常会仿译成为汉语的显性条件句,并增添其他连词和人称代词(如仿译英文斜体部分)。若知道做隐化处理,才有可能回到地道的汉语原文。如下面的例句:

(8)a With my present level (of French), I can not read French newspaper easily.

(8)b 以我现在的(法语)水平,我还不能轻松地读法文报纸。

(8)c 我的法语,看报很吃力。

(9)a If you take this medicine, your illness will surely be cured.

(9)b 你要是服用这个药,你的病肯定能治好。

(9)c 这药吃了准好。

(10)a You needn't care about the affairs in the home.

(10)b 你不必担心家里的事。

(10)c 家里的事,不用你管。

对于(8)a,(8)b 是仿译,(8)c 做隐化处理,以汉语的主题句译之。对于(9)a,(9)b 是仿译,(9)c 做隐化处理,以汉语隐性条件句译之,并省去人称代词。对于(10)a,(10)b 是仿译,(10)c 以汉语的宾语前置主题句译之。由此可见,英译汉时做一定的隐化处理,译文会更地道,也比较简洁。仿译则将外语的表达法引入汉语,对于汉语来说,已经在形式上显化了,即本可以隐化表达的意思外在地显示出来了。这一类情况在英译汉中表现得非常明显和频繁。

除了上述有代表性的例句外,我们对汉英对应语料库中 80 多万汉字的汉译英语料(汉语为原文)和近 80 万汉字的英译汉语料(汉语为译文)加以考察,发现以下表示时间的连词词组的用法在汉语原文和汉语译文中出现的频率达不相同:

在汉译英语料(汉语原文)中,在……时 20 处,在……时候 99 处,当……时 6 处,当……时候 23 处,共 148 处;

而在英译汉语料(汉语译文)中,在……时 279 处,在……时候 140 处,当……时 80 处,当……时候 18 处;共 527 处。

表示假设条件的连词用法在汉语原文和汉语译文中也相差很远:

"如果"句:汉语原文中仅见44句,汉语译文中多达393句。"要是"句,汉语原文语料中有81处,汉语译文语料中出现139处。

还有表示结果的连词"以致"在汉语原文和译文中的使用差别很大:

"以致"句:上述汉译英语料中未见一句(0),英译汉语料中见到20句。

这些数据说明,与原文语言相比,译文语言确实有表达上的特点,是无形中受到的源语言影响,同时也说明翻译中采取隐化策略之必要。

此外,从上述分析可以看出,隐化和显化与翻译方向有一定关系,至少在不同翻译方向上,隐化和显化程度是不同的。结合 Øverås(1998)的考察,我们不妨假设:由形式化程度较高(或形态繁复)的语言翻译成形式化程度较低的语言,如由挪威语翻译成英语,或由英语翻译成汉语,显化现象发生递减,而隐化现象发生递增。若是相反的翻译方向,则显化递增,隐化递减。这一假设有待于进一步的语料库验证。

9.4　广义上的翻译显化现象

作为一种翻译现象,显化(以及隐化)不应只是狭义地指语言衔接形式上的变化,还应包括意义上的显化转换,即在译文中增添了有助于译文读者理解的显化表达,或者说将原文隐含的信息显化于译文中,使意思更明确,逻辑更清楚。这也是翻译特有的现象。

许国璋论述翻译时,曾提出对哲学著作的翻译应采取"通译"、"切译"或"阐译"、"释译"方法(1991:248,262)。他力主译文必须通脱、醒豁,译出词的文化史涵义。他特地翻译罗素《西方哲学史》以为其"阐译"示范。他解释说:*feudal* 不译"封建",译"拥据领地(之诸侯)";*anarchy* 不译"无政府"(当时无中央政府),而译"诸侯纷争";*adventure* 不译"冒险",或译"猎奇于远方",或译"探无涯之知",视上下文而定,此而不济,则作"释译"以为助。例如 *antiquity* 不译"古代",译"希腊罗马"。又如 *fame and beauty*,在文艺复兴时期有特殊涵义,今试译"享盛誉于邦国,创文艺之美";*asceticism* 译"绝欲弃俗,攻苦食淡之说教";*prudential arguments* 译"保身家保名誉之考虑"(1991:248)。

这些并不是 Øverås(1998)所讨论的衔接手段上的显化,而是意义上的显化,但这也是翻译中常见的情形,同意值得探讨。翻译家的责任不仅在于通过语言作翻译介绍,他还需要为自己的文化引进相应的概念系统,为读者更好地理解作品提供背景知识。我们再看两例许氏译文:

(11) Gunpowder strengthened central government at the expense of feudal

nobility.

(11)a 火药用于争战，中央政府因之以强，拥据领地之公候因之而弱。

(11)b 火药消灭了封建贵族而巩固了中央集权政治。(何、李译，p. 589)

(12) When Constantinople, the last survival of antiquity, was captured by the Turks, Greek refugees in Italy were welcomed by humanists. Vasco da Gama and Columbus enlarged the world, and Copernicus enlarged the heavens.

(12)a [十五世纪中]君士坦丁堡为土耳其所占。自希腊罗马之衰，古文化残存于世，仅君士坦丁堡一地。既陷，寓居于君士坦丁堡之希腊学人，相率流亡意大利，意大利倾慕希腊人文之学者，迎为上宾。达迦玛绕好望角抵印度，哥伦布西行发现新地，而世界为之扩大。哥白尼立日中心之说，而宇宙为之扩大。

(12)b 当君士坦丁堡，这个古代最后的残余，被土耳其人攻陷后，逃往意大利的希腊难民曾受到人文学者的欢迎。瓦斯寇·达·伽马和哥伦布扩大了世界，而哥白尼扩大了天界。(何、李译，p. 589)

对照原文，译文(11)a 和(12)a 在意思上充分显化，原文隐含的或在其文化上不言自明的内容，已为译者以加词、变通、阐释等方法在译文中显化出来，形成一种阐译，比直译(如例(11)b 和例(12)b 的译文)甚至比直接读原文更清晰、易懂。例如读例(11)b，读者不容易弄懂火药怎样消灭封建贵族而巩固中央集权，(11)a 添加“用于争战”，将原文隐含的因果关系之意显化，就明白易懂了。例(12)更是如此。相比于 b，a 的下划线部分添加了不少文化背景上的解释，如“最后的残余”解释为“古文化残存于世，仅君士坦丁堡一地”，使人明白残余指文化之残存；达迦玛、哥伦布、哥白尼如何使世界扩大？a 将其背景显化在译文中：“绕好望角抵印度”，“西行发现新地”，“立日中心之说”，读者于是明白。这种显化现象无疑与译者的翻译主张相关：“翻译目的，在于便利不懂外文之读者，如不懂外文之读者读之不懂，翻译者不能说尽到责任。”(许国璋 1991：248)

9.5 小结

翻译中的隐和显，由多种因素造成，如 1)语言因素：隐、显现象的发生和隐、显程度可能与语言的形式化程度或形态形式和翻译方向相关。例如汉语同英语等印欧语言相比，其词汇、语法衔接上的显化度都比较低，汉译英时会比英译汉时更多地呈现形式上的“显化”，而英译汉时则呈现更多的模仿式的“显化”。2)译者因素：

负责的译者在穿梭于两种语言之间进行协调时，总会尽量减少信息传输过程中的损耗和丢失，便于读者的理解和吸收。但另一方面，有些译者因其能力上的不足，在翻译中对原文亦步亦趋地仿译，也会导致译文过度“显化”。3）社会文化因素：翻译涉及的两种语言在社会和文化上的差距越大，可能越需要解释性的“显化”。4）文本因素：小说文本、科学文本、政论文本、新闻文本、应用文文本等在翻译中的显、隐程度可能有别，本章未及论述。

第10章

翻译汉语的词汇简化与显化

本章通过“通用汉英对应语料库”中翻译汉语与原创汉语语料的分析和对比，考察翻译汉语区别于汉语原创语料的词汇使用特征。语料库研究表明：无论是文学还是非文学语料，翻译汉语与原创汉语相比，具有词语变化度偏低、词汇密度偏低、虚词显化、指代方式显化、常用词频率增加等特征。本章还就这些特征探讨了“翻译共性”问题。

10.1 翻译语言的特征与翻译共性

翻译涉及原语和译入语两种语言，所产译文必然与这两种语言及其文化相关。译文再“异化”，也不可能全部照搬原语而不顾译入语的语言规则规范；译文再“归化”，也总会与译入语存在某些差异。这些差异不会以明显不符合目的语语法规则的形式出现，而往往以多少有些异常的形式偏离语言使用的规范（Toury 1980：226）。我们把译文语言与译入语语言规范的差异称为“翻译的特征”（features of translation），并由此探讨“翻译共性”（translation universals）。

对翻译自身特征的系统研究仅有十几年历史。Baker（1993）、Kenny（2001）、Laviosa（2002）、Olohan（2004）等先后采用语料库语言学的方法来描述翻译英语的特征。他们据此所揭示的翻译共性主要是：1）简化（simplification），指“译者无意识地简化语言或信息或两者兼而有之”（Baker 1996：176）。换言之，即翻译文本比原创文本的语言更简单，例如用词范围缩小、常用词增多、句型简化等。2）显化（explicitation），指对原语中暗含的、需从上下文中推导出的信息在译语中予以明示（Vinay & Darbelnet 1958/1995）。显化最主要的表现是：扩展句子和语篇；增添修饰词、限定词和连接词，增添额外信息，增添解释语，代词明示等。3）范化（normalization），指译者消除原文独特的文本特征，用译入语中典型、规范的方式来生产译文，而且译文可能比译入语的原创文本更加“规范”。主要表现为：特色词减少，常用词增多，标点符号更规范，繁琐、奇异的句子减少，逻辑性更强等。除上述三个翻译共性外，个别研究者还提出了删减原文中的重复、译入语特征夸大化等

特征。概言之，翻译特征可描述为语言更简单、易读性更强（简化）；歧义更少、表达更清楚（显化）；更符合译入语规范（范化）；同时，翻译文本的独创性可能降低，译文偏向于译语文化中的可接受性。学者们倾向于将这些翻译特征逐步论证为翻译共性，这一点自然需要更多的双语翻译语料的验证。

上述翻译特征研究主要在英语为主的欧洲语言间的翻译中进行。我国的语料库翻译学经过前几年的介绍阶段（廖七一 2000；张美芳 2002；柯飞 2002；胡显耀 2005 等）后，也开始进行基于汉英对应语料库的相关研究。本章即利用北京外国语大学"汉英对应语料库"[①]对翻译汉语（包括文学和非文学语料）不同于原创汉语（指汉语原创语料）的词语使用特征进行描述和分析。需说明的是，本章的出发点是采用语料库方法来验证翻译汉语是否存在翻译共性或区别性特征的假设。本章的比较是在同为汉语的翻译和非翻译的类比语料库之间进行，与传统对比语言学在语言间的比较有区别。

10.2 翻译汉语的词汇特征

10.2.1 词语变化度

词语变化度是指相同长度的语料中不同词语的数量，可以作为语料难易程度的衡量因素之一，以语料库的类符形符比（type/token ratio，TTR）来衡量。类符形符比指特定语料中类符（type）与形符（token）的比率。类符是语料库中不同的词语；形符是所有的词形。TTR 值越大说明该文本所使用的不同词汇量越大，反之则不同词汇越少。但由于一种语言在特定时期的词汇量是有限的，语料库容量不断扩大，形符数持续增加，但类符数却不一定会增加；语料库容量越大，形符类符比反而会越来越小，因此不同容量的语料库的形符类符比不具备可比性。故一般用标准类符形符比（STTR）来衡量语料库的词汇变化，即按一定的长度分批计算文本的类符形符比，再求均值类符形符比反映文本的信息量的大小。用 Wordsmith Tool（WS）可自动计算 TTR 和 STTR。表 10.1 为使用 WS4.0 统计的语料库基本信息（STTR 计算基数为 100）：

---

① 语料库容量为 3 000 万字词，由翻译、百科、专科和对译语句等四个子库构成。其中翻译文本库容量为 2 000 万字词，英译汉占 60%；汉译英占 40%，分别包含文学和非文学语料。全部语料进行了句对齐和词性标注。文学语料以小说为主。本章主要使用四个子库的语料：1）汉语文学原创语料；2）汉语文学翻译语料；3）汉语非文学原创语料；4）汉语非文学翻译语料。

表 10.1 各子语料库的词语变化度比较①

	文学		非文学	
	原创汉语	翻译汉语	原创汉语	翻译汉语
形符数(token)	403 418	564 390	302 303	460 887
类符数(type)	14 771	16 340	10 151	17 111
TTR	3.66	2.9	3.4	3.77
STTR	71.84	70.79	66.95	70.87
STTR 标准差	29.00	28.28	32.34	30.17
平均词长(字数)	1.28	1.32	1.55	1.46
词长标准差	1	0.55	0.73	0.67

表 10.1 列举了反映词语变化的各项指标。总体特征是:文学语料与非文学语料的表现呈相反趋势——汉语文学语料中翻译汉语的 TTR 和 STTR 小于原创汉语;而非文学语料库翻译汉语 TTR 与 STTR 大于原创汉语。这意味着:文学语料的词语变化趋势支持"简略化"的假设,即汉语文学翻译语料显示使用较少的词汇,而原创文学倾向于使用较多的词汇。词语变化特征表明:汉语文学翻译语料支持"简化"共性假设(Laviosa 1998,2002;胡显耀 2007),而汉语非文学翻译语料似乎不支持这一假设。关于后者,我们分析是所用语料库中非文学原创汉语语料的选材稍显单一,主要为政府文件类语料(共 199 个文件);而翻译汉语则涉及科技、经济、法律、医疗等各个领域(共 400 个文件)。两个语料库的规模和选材有差距,因而可比性略差。尽管文学语料的词语变化度支持翻译简化的假设,但由于语料库之间的类符形符比的差距并不特别悬殊,这些差异还可能受语料库工具的精度和语料取样的范围的影响,所以它所能说明词语简化的程度有限。

10.2.2 词类频率分布

本章所使用的四个语料库都做过词性赋码,因此得以对各子语料库的各类词性进行全面的统计分析。我们用 PowerGrep3.0 检索了语料库中全部 43 个 POS

① WS 统计的数据与采用 PowerGrep3.0 等文本检索工具检索的数据有一些差异。这可能是由于 WS 内部对汉字形符计算的方式有所不同。本章对相同数据采取完全相同的工具以确保词表数据具有可比性。

标记的频次，并依据各子语料库的实际形符标准化了各种词类的每百万词频率[①]（见表 10.2）。为了分别统计实词和虚词的频率，本章在表中用"—"表示虚词。鉴于代词属实词或虚词还存在争议，及代词在翻译汉语中的特殊表现，本章用"*"单列表示。以下分别从词汇密度、虚词形符比、代词形符比三个方面具体分析翻译汉语的词类频率分布特征。

表 10.2　各子语料库词类频次比较(每百万词)

	文　学		非文学	
	原创汉语	翻译汉语	原创汉语	翻译汉语
动词	206 851	200 358	212 206	203 975
名词	175 609	151 910	240 423	232 432
副词	85 898	75 695	52 778	57 232
形容词	44 232	41 687	41 843	40 492
数词	30 895	29 768	35 130	34 966
量词	25 297	23 942	17 319	20 427
代词	*74 307	*101 141	*41 856	*53 154
助词	—70 697	—83 437	—75 005	—84 320
介词	—26 064	—37 692	—37 545	—45 807
连词	—17 359	—21 950	—28 944	—33 594
语气词	—14 450	—11 876	—4 044	—2 673
叹词	—1 808	—1 710	—93	—138
合计	773 467	781 166	787 188	809 210

10.2.2.1　词汇密度

词汇密度(lexical density)有两种计算方法。一种是以 TTR 值作为词汇密度（杨惠中 2002:168）。一般认为，这种词汇密度高于 60%～70%为密度较高，低于 40%～50%为密度较低。这种词汇密度将词语的变化性等同于词汇密度，但如上所述，TTR 对语料库容量和选材十分敏感，不能真实反映词语变化度。另一方面，STTR 也不能完全反映篇章的信息量，原因是 STTR 统计的类符包括实义词和功

① 本节的某一词类包括该 POS 的全部赋码，例如：形容词包括 POS=a，ad，ag，an；名词包括 POS=n，ng，nr，ns，nt，nx，nz 等。全部词性赋码标记集(Tagset)参看《北大汉语文本词性标注标记集》。

能词，过度修饰的篇章由于功能词的增加也可能提高 STTR 值，但并不意味着该语篇信息量的增加。

另一种是 J. Ure(1971)和 Michael Stubbs(1986)提出的方法，即统计实词形符在总形符数中的比例。英语的实义词(lexical word 或 content word)指具有稳定词义的词语，包括名词、动词、形容词和副词等词类。与之相对的是功能词(function word)，指不具备稳定词义或意义模糊而主要起语法功能作用的词语，包括介词、连词、冠词、助动词等词类。汉语词类划分和实词与虚词的区分至今还存在争议[①]。本章用前一种方法词汇密度来考察和比较翻译汉语与原创汉语在信息量上的差异和难易度，同时也采取第二种方法，即在具体统计中把名词、动词、形容词、副词、数词、量词等六个“具有稳定词义”的词类作为实词。将汉语中起语法功能的助词、介词、连词、语气词、叹词等视为虚词。而“半虚词”代词(王力)的意义必须依靠与其同指的名词才能确定，本章认为它不应算作具有“稳定的词义”。但是，我们在对各词类频率及比例的统计中发现，汉语文学翻译中代词的使用与非翻译语料具有显著的区别，故本章对代词单列统计并专门讨论。表 10.3 中列举了三类词的频率在各子语料库中的比例。

表 10.3 各子语料库词汇密度对比

	文 学		非文学	
	原创汉语	翻译汉语	原创汉语	翻译汉语
实词	61.40%	57.30%	64.50%	63.10%
代词	7.43%	10.11%	4.19%	5.32%
虚词	13.04%	15.67%	14.56%	16.65%

① 汉语实词与虚词和各种词类的划分存在不同的划分标准和类型。如《马氏文通》把汉语分为实词和虚词两大类，“凡字有事理可解者，曰实字；无解而惟以助实字之情态者，曰虚字。”虚字分“介字、连字、助字、叹字”四类。黎锦熙《新著国文语法》将词分五大类：实体词、述说词、区别词、关系词、情态词。其中虚词是指关系词和情态词。王力《中国现代语法》认为词分为理解成分和语法成分，理解成分即实词包括：名词、数词、形容词、动词；语法成分包括半实词(副词)、半虚词(代词、系词)和虚词(联结词和语气词)。吕叔湘和朱德熙《语法修辞讲话》则把名词、动词和形容词归为实词，把代词、副词、连词、语气词、象声词等归为虚词。可见，争议最大的词类是副词和代词，究竟这两类词属于虚词还是实词，是汉语语言学家争论不休的问题之一。本章无意讨论汉语词类划分的问题，但由于在研究翻译文本与非翻译文本差异性时，不可避免地涉及汉语的词类问题，因此本章综合各家意见，主要采取了王力、吕叔湘与朱德熙先生看法，将名词、动词、形容词三类词归为实词；而将副词、代词、介词、连词、助词和叹词等归为虚词。

（续表）

	文学		非文学	
	原创汉语	翻译汉语	原创汉语	翻译汉语
标点	17.00%	15.80%	15.23%	13.27%
其他	1.12%	1.06%	1.51%	1.62%

表 10.3 中实词的比例即本章所说的词汇密度。可以看到，无论是文学语料还是非文学语料在词汇密度上表现趋于一致：翻译汉语的词汇密度低于原创汉语，即相同长度的翻译汉语语料中使用的实词少于原创汉语，或换言之，相同长度的翻译汉语语料中可能使用了更多的功能词，词汇密度低反映信息量和文本难易度相对降低。因此，词汇密度在各子语料库中的表现支持翻译文本的“简略化”假设——即汉语翻译文本具有通过降低实词所提供的信息量来降低译文的难度，从而提高其可接受性的总体趋势。

10.2.2.2 虚词显化

上文考察了实词的比例，那么虚词在翻译汉语中表现如何？从表 10.3 中可以看到，汉语文学和非文学语料具有共同趋势：翻译汉语的虚词比例高于原创汉语。也就是说，翻译汉语语料中起语法功能的介词、连词、助词等词类使用频率比原创汉语文本高，我们把这种现象称为“虚词显化”（或语法显化）。由于汉语总体而言趋向语法标记内隐（意合），而翻译汉语在语法标记上则出现了外显（形合）的特征，我们认为，这一特征支持翻译汉语显化共性的假设。不过，虚词在语料库中的具体表现如何？造成翻译汉语虚词显化的原因是什么呢？为了回答这些问题，我们对语料库中具体词类进行了更深入的分析。

图 10.1 是文学语料中原创汉语和翻译汉语的各主要词类每百万次频率比较。从中可以更直观地看到文学语料中：1）原创汉语的全部实词频率均高于翻译汉语，其中以名词频率差异较为显著。2）原创汉语的全部虚词频率均低于翻译汉语，其中助词、介词、连词均有较大差异。3）原创汉语的代词使用频率明显低于翻译汉语。

图 10.2 是非文学语料中原创汉语和翻译汉语的主要词类频率比较。从该图可以看到汉语文学语料与非文学语料的差异：总体而言，文学语料的实词频率更低而虚词频率更高；文学语料中频次最高的词类是动词，而非文学语料中频次最高的是名词。这种差异可以从文学与非文学语料的不同功能来解释：文学作品主要供人消遣娱乐，非文学作品的主要功能是提供信息，因而文学语料通过降低词汇密度，增加语法明晰度来提高自身的可读性。除此以外，非文学语料中，原创与翻译汉语体现出与文学语料类似的趋势，即原创汉语的实词频率高于翻译汉语，而虚词

频率低于翻译汉语。但从虚词频次差异程度来看，非文学翻译语料不如文学翻译语料明显(见图 10.1 和图 10.2)上述比较说明虚词显化可以作为文学与非文学翻译汉语语料的共同特征。这一特征使我们对汉语翻译文本的语言特征有了新的认识。

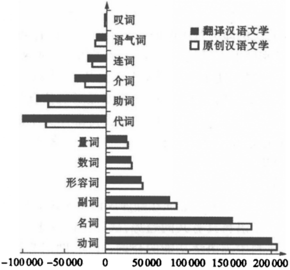


图 10.1　原创汉语和翻译汉语文学语料词类频率对比

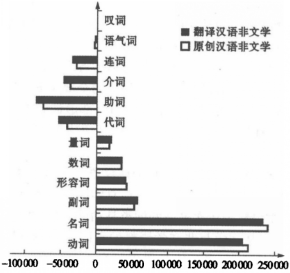


图 10.2　原创汉语和翻译汉语非文学语料词类频率对比

10.2.2.3　指代显化

如上所述，本章把代词作为“半虚词”单列讨论，这是由于代词在翻译汉语中出现了与原创汉语显著不同的特征。汉语代词除具备和名词相同的指别(deixis)的功能外，还具有照应(anaphora)功能，即与其他语言单位构成同指关系的功能。王力(1989)把代词归入“半虚词”，就是指代词的照应功能。与英语及其他印欧语言相比，汉语代词的类型较少，使用频率较低(刘宓庆 1998：445-476；范仲英 1997：148-151)。汉语人称代词形式上没有主格宾格之分，代词所有格与英语的物主代词相比使用得更少。汉语常规的指代方式主要以“名词复现”和“零代词”为主，显性人称代词的使用频率一般较低。上述特征可以说是原创汉语的“常规”，但从语料库代词频率统计来看，翻译汉语中代词显然偏离了这一常规。从图 10.1 和图 10.2 汉语学和非文学语料的代词频率对比来看，翻译汉语的代词频率明显高于原创汉语。这说明翻译汉语倾向于将指代关系显化。

10.2.3　词表与范化

词表(wordlist)指按照类符的频率高低列出的类符表，也叫词频表。我们用 PowerGrep 工具制作了四个子语料库的词表(见表 10.4；限于篇幅本章仅列前十位)。该词表列出所有类符的频率(frequency)和在总词频数中所占的百分比。初步比较之下，各子语料库的词表开头部分非常相似：除了汉语非文学原创语料库外，其他三个语料库词表的前 30 位都是以“的、在、和、了、是、一、为、有、不、我、他”等单字词构成的。虽然各子语料库的具体顺序略有差别，但基本上大同小异。这些单字词的高频现象是汉语词表的普遍现象①。各子语料库前 30 位的高频词几乎都是助词、介词、连词和代词等起语法功能和修辞作用的虚词。现代汉语所有的结构助词“的、地、得”和动态助词“着、了、过”均出现在前 30 位中，其中结构助词“的”在所有语料库中皆排在第一位，各子语料库中频率最高的动词为“是”，介词为“在”，数词为“一”，副词为“不”，名词为“人”。

表 10.4　各子语料库词表前十位

	汉语文学原创			汉语文学翻译			汉语非文学原创			汉语非文学翻译	
	频率	%		频率	%		频率	%		频率	%
的	22041	4.78%	的	38017	5.97%	的	21425	6.71%	的	37999	7.73%
了	12814	2.78%	了	15476	2.43%	是	5589	1.75%	在	7964	1.62%
不	8182	1.77%	他	15446	2.43%	和	5306	1.66%	是	7852	1.60%

① 汉语非文学原创语料库选材稍显单一(以政论文为主)，故其词表特征表现略特殊。

（续表）

	汉语文学原创			汉语文学翻译			汉语非文学原创			汉语非文学翻译	
	频率	%		频率	%		频率	%		频率	%
是	7920	1.72%	我	11859	1.86%	了	4290	1.34%	了	5626	1.15%
他	7666	1.66%	是	10459	1.64%	在	3949	1.24%	一	5299	1.08%
一	6322	1.37%	在	10232	1.61%	中国	3787	1.19%	和	5213	1.06%
我	5052	1.10%	一	8699	1.37%	不	3189	1.00%	不	3575	0.73%
在	4710	1.02%	不	7395	1.16%	我们	2720	0.85%	有	3096	0.63%
着	4200	0.91%	你	6094	0.96%	有	2419	0.76%	人	3007	0.61%
你	4122	0.89%	她	5713	0.90%	人民	2275	0.71%	他	2987	0.61%

翻译汉语语料与其他原创汉语词表的显著区别在于：1)汉语文学翻译词表前 30 位词语的频次比例之和为 32.27%，明显高于其他汉语语料库（文学原创 29.99%，非文学原创 25.33%和非文学翻译 24.35%）。这意味着汉语文学翻译语料中词频最高的 30 个词重复使用的频率更高。2)文学翻译词表前十位中，出现了四个人称代词（"他、我、你、她"），而文学原创中只有三个（"他、我、你"），非文学原创和非文学翻译各仅一个人称代词。这说明，文学翻译中主要人称代词的使用频率高于其他所有语料库，这为"指代显化"提供了另一证据。

Laviosa(1998)用高频词（即最常用的词语）的范围来考察英语翻译叙事文的词语使用的模式。她提出并验证了翻译英语中高频词所占比率高于非翻译英语，词频最高的表头（list head）占更高的比率，但词目（lemma）数量却更少（1998：6-8）。汉语翻译语料中的高频词是否也有同样的特点呢？我们分析了各词表中不同词频段的高频词数量和频率。

图 10.3 显示的是各子语料库词表中词频高于 1%，0.5%～1%，0.4%～0.5%，0.3%～0.4%，0.2%～0.3%，0.1%～0.2%的词语数量。可以看到各子语料库中词频高于 0.1%的高频词数量分布。汉语文学语料和非文学语料中，翻译汉语高频词数量特征是一致的——翻译汉语中高频词（常用词）的数量少于原创汉语。这表明翻译汉语倾向于重复使用较少数量的常用词。这一特征与 Laviosa(1998)所说的"英语翻译小说词汇模式"是一致的。

另外，这些高频词的使用频率与各子语料库总词频的比例有何差异？我们分别统计了各子语料库词表中排名前 10 位、30 位、50 位和 100 位的类符在总词频中的比例（图 10.4）。该图可作如下分析：1)各阶段高频词频率汉语文学语料均高于非文学语料，这说明文学文本与非文学文本相比，常用词更多。2)文学语料中，翻译汉语各阶段高频词频率均高于原创汉语，即文学翻译文本使用更多常用词。

3)非文学语料中,翻译汉语各阶段高频词频率均低于原创汉语,即非文学翻译文本常用词更少,但差距不明显。

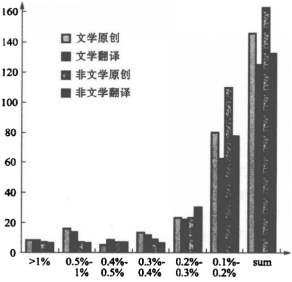


图 10.3　各语料库高频词数量比较

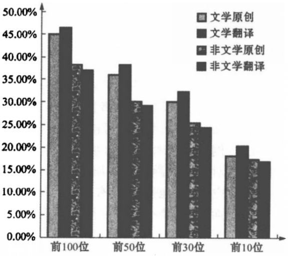


图 10.4　各语料库词表高频词比例

上述分析说明:文学翻译语料中使用的常用词汇数量更少,而使用频率更高,但非文学翻译语料似未呈现同样趋势。换言之,在进行英汉文学翻译时,译者群体有意或无意地使用重复使用了较少数量的常用词语。我们认为这一特征支持翻译

汉语简化和范化假设——常用词增加，文本独特性减少，文本难度降低，可接受性提高。不过，值得注意的是，非文学翻译语料只是部分支持这一假设。如前所述，这可能是受语料选材的影响，希望以后的研究能更多关注非文学翻译的特征。

10.3 小结

从本研究来看，我们可以对翻译汉语的词汇特征进行如下概括：汉语文学翻译语料与文学原创汉语相比，表现出明显的简化、显化和范化特征——1)类符形符比较低，说明文学翻译的词语变化度较低；2)词汇密度较低，表明文学翻译实词数量和频率较低，故实词数量和频率所反映的信息量大小和文本难度也就相应降低；3)文学翻译中，各种虚词形符频率和代词频率均高于原创汉语，即文学翻译出现了虚词和代词显化特征；4)文学翻译中常用词数量更少而频率更高，而非文学翻译语料与同类原创汉语语料相比，除类符形符比和高频词外，也出现了词汇密度降低和虚词代词显化特征。因此，我们至少可以把实词简略化和虚词、代词显化作为翻译汉语共同的词汇特征。

上述研究是在原创汉语与翻译汉语所构成的类比语料基础上进行的，它揭示的是同为汉语的翻译文本与非翻译文本的差异性特征。这一点是本章和基于对比语言学的翻译研究的区别。以语料库为基础的研究为描写大量文本的共同趋势提供了很大便利。不过，词汇的简化与显化特征只是翻译汉语的总体特征，并不意味着所有汉语翻译文本都会出现同样的简化和显化。个别译者在处理具体译文时既可能遵从也可能违背这种趋势。不过个别译作的倾向不会改变特定时期汉语文学翻译语料的总体特征。此外，实词简化也不能简单等同于汉语文学翻译语料阅读难度更低，原因是异域文本对译入语文化的读者往往是陌生的。尽管译者通过降低实词密度，提高语法明晰度的方法来增加其可接受性，但限于语际翻译本身的性质，这种陌生感总会存在的(例如人名地名音译，物质、文化和事件名词的差异等)。

翻译的特征或翻译共性对于探索翻译的本质，了解语言交往时的相互作用和研究译入语在表达语言文化空缺时的创造力具有重要的意义。而以语言对比为基础的传统翻译学要么对此视而不见，要么把偏离译入语规范的翻译语言特征笼统地贬斥为“翻译腔”。从这个意义上说，基于语料库的描写翻译学做出了具有开创价值的贡献。当然，我们也应该看到，尽管语料库翻译研究为我们提供了对大量真实翻译文本的客观描述，但方法本身并不提供对翻译现象的解释。对翻译语言的特征或翻译共性的认识需要在更广泛、更细致的描写的基础上不断地加深。

第 11 章

汉译文学作品的语法显化

本章基于语料库对汉语翻译小说中语法标记词的考察发现：汉语翻译小说中各类语法标记词的使用频率明显高于非翻译小说；翻译小说定语标记明显增多，而时体标记减少；一些语法标记词的增减在客观上起到了口语化或简化的效果。“语法标记显化”是汉语翻译小说的重要特征之一。该特征一方面体现了翻译小说语法的明确程度提高，旨在加强译本的可接受性；而另一方面却使汉语译文在一定程度上偏离了现代汉语规范，呈现出句法形式的陌生化或异化倾向。

11.1 翻译共性与语内显化

翻译文本中的“显化”(explicitation)现象近年来在国内外学者中引起了较广泛的注意(Øverås 1998；Olohan & Baker 2000；柯飞 2005；胡开宝，朱一凡 2008；黄立波 2008 等)。但是，从国内外学者对显化概念的使用来看，其间存在较大理解上的分歧。因此，明确厘定“显化”概念是研究“翻译共性”(translational universals)或“翻译特征”(features of translation)的前提之一。

首先，我们知道：翻译共性的研究背景主要是由埃及裔英国学者 Mona Baker 所创的语料库翻译研究。何谓“翻译共性”？Baker 曾明确定义：“译文而非原文中的典型的语言特征，这些特征与翻译过程涉及的特定语言对无关”(Baker，1993：243)。她的学生 Kenny 也说：“普遍性是一种总体的倾向，与译者、语言、文体和时代无关”(2001：52)。Olohan & Baker(2000)、Kenny(2001)、Laviosa(2002)、Olohan(2004)等人的语料库翻译研究都是围绕着这一概念展开的，即通过同一语言内翻译与非翻译(或原生/原创)文本的比较，寻找翻译文本共同的特征。

第二，何谓“显化”？该概念由法国人 Vinay 和 Darbelnet 与 1958 年提出，指“对原语中暗含的，但可以从上下文中推导出的信息在译语中加以明示”(1958：8)，与之相对的是“隐化”(implicitation)。实际上，这对概念最初是作为翻译技巧与“增补”和“删减”相提并论的(见 Klaudy 2004：80-84)。20 世纪 80 年代，Blum-Kulka 系统研究了显化现象，提出了著名的“显化假设”，即显化程度提高可能是翻

译文本和二语习得等语言中介活动中具普遍性的策略(1986:19;21)。Blum-Kulka认为显化是翻译本身固有的特征,与译者对原文的理解过程有关,而与特定的语言对无关。可见,显化假设与作为翻译技巧的显化内涵不同。这里的显化才是Baker所说的翻译共性之一。

但从近期的一些论文来看,国内同仁对显化的理解与Blum-Kulka和Baker等人存在着差异。由于语言形式化程度、文体、译者和文化差异等因素的存在,翻译的具体过程中确实普遍存在对原文语义和句法形式的显或隐(柯飞2005)。但从实例来看,翻译中的"显隐"与翻译技巧"增减"似无本质区别,都是根据译入语的语法规范增添或删减语义内容或语法标记,且显隐的程度随各种因素的改变而改变。这显然与Baker等人所说的与特定语言对无关的"翻译共性"的概念不一致。

黄立波(2008)注意到了国内学者理解的显化与Baker等人的区别,因而提出了"语际显化"和"语内(类比)显化"的分类。"语际显化"指译文中对原文隐含信息的明示;"语内显化"则是翻译文本相对同为译语的非翻译文本的显化程度提高。国内学者对显化的研究主要是对"语际显化"的研究,即英汉互译过程中的显化(和隐化)。

要厘清"显化"概念的所指,消除研究上的混淆,我们不妨从比较对象、显化内容、强制性和影响因素四个层面加以澄清:

表11.1 "显化"概念分析

<table>
<tr><th></th><th>比较对象</th><th>显化内容</th><th>强制性</th><th>主要影响因素</th></tr>
<tr><td rowspan="2">语际显化</td><td rowspan="2">原语文本/译语文本</td><td>语义显化</td><td>非强制性</td><td>译者社会</td></tr>
<tr><td>句法显化</td><td>强制性/非强制性</td><td>语言差异</td></tr>
<tr><td rowspan="2">语内显化</td><td rowspan="2">非翻译文本/翻译文本</td><td>—</td><td>—</td><td>—</td></tr>
<tr><td>句法显化</td><td>强制性</td><td>翻译认知</td></tr>
</table>

根据比较的对象,显化可分为"语际显化"和"语内显化"。前者属于对比语言学范畴,而后者指同一语言内翻译与非翻译文本的差异。从内容上讲,显化可以是对原文暗含语义的明示,也可以是对语言形式本身的明示,也即"语义显化"和"句法显化"。语义显化受译者、文体、读者和社会因素等的影响;语义是否显化和显化的程度随译者和社会需求的改变而改变,因而它是非强制性的。在语际显化中,句法显化取决于原语和译语的句法形式化程度的差异和翻译的方向。它可以是强制性的(如汉译英),也可以是非强制性的(如英译汉)。总之,语际显化和隐化本质上是一对凸显或删减语义和语法标记的翻译方法,各种因素制约着译文显隐的程度,它与翻译共性中的"显化"不同质。

Baker等人讨论的"显化"实际上是指翻译与非翻译文本之间显化程度的差异,即"语内显化"。在语内显化中,由于翻译和非翻译文本不存在内容上的关联,故也无所谓语义显化;两者的主要区别体现于文本特征和语言形式的差异,也就是句法显化。Baker等人研究的目的就是为了验证"显化、简化和规范化等是翻译文本的普遍特征"这一假设。这些普遍特征是由翻译和认知因素决定的,与语言、文体、译者和社会文化无关。简言之,"显化"是翻译文本的本质特征之一。当然,在对各种语言的翻译文本进行全面的实证研究之前,语内显化仅仅是一种有待验证的假设,这就是Blum-Kulka(1986)提出的"显化假设"。

值得指出的是,在语料库翻译研究中,显化是一个总体性概念,即翻译文本总体与非翻译文本总体相比较而言。两者须具备语料库语言学上的代表性和可比性。这个概念与具体翻译过程中的语义或句法显化(即上文所说的语际显化)有根本区别。后者是一个技术性、个别性和非强制性的问题;而前者是一种普遍趋势。本文所讨论的"语法标记显化"就是指汉语翻译文本区别于汉语非翻译文本的总体特征。

11.2 语料库与翻译小说的"形合度"

语料库翻译研究的主要目标是用语料库语言学的工具和方法来描写真实的翻译现象,以探寻翻译自身固有的语言特征,或寻找翻译的普遍性质(详见胡显耀:2005,2008)。这种方法与基于对比语言学的翻译研究的区别在于:它要描写的是翻译作为一种语言变体的总体特征;而后者以实用为主要目的,旨在发现和分析语际转换的方法和技巧。

语料库翻译研究须具备一个前提:建立具代表性的翻译语料库和与之有可比性的非翻译语料库。为此,Baker及其学生Laviosa建立了第一个只收入翻译语料(不限原语种类)的"翻译英语语料库"(TEC)。我国目前已有的语料库翻译研究中,绝大多数并非通过可比语料库(comparable corpus)途径,而是以平行语料库(parallel corpus)为主。为了研究汉语翻译文本的总体特征,建立能够代表汉语翻译文本的可比语料库势在必行。

为此,我们建立了一个翻译语料库——"当代汉语翻译小说语料库"(CCTFC)。该语料库收入了译自英、美、法、俄、德、日等11个国家八种语言的56部汉语翻译小说。语料1980年至2000年间出版的长篇翻译小说。每部作品以取样方式收入20 000词,目前容量130万词。该语料库是国内第一个现代汉语单语翻译语料库,即只收入翻译小说语料(不限原语种类)。CCTFC进行了词语切分、词性标注、句子和段落等多层次标注。本章将围绕CCTFC以及作为类比语料库的"兰开斯特

现代汉语语料库"(LCMC)①的小说子库 LCMC(N)展开。

如上节所述,语内显化主要表现为句法显化。各种语言的句法形式化程度是不同的。以英、汉语为例,形式化问题在我国一般被称为"形合"与"意合"。"形合"指用词语、形态变化、语法等手段来连接短语和句子;"意合"则指借助语序来体现词句之间的意义或逻辑关系。英语形式化(或形合)程度较高:如大量借助语言形态变化,频繁使用关系词、连接词、介词等连接词句和篇章等。而汉语形式化程度较低:几乎没有词形变化②;缺少与英语类似的关系词;连词和介词都不活跃。不过,现代汉语也大量采用助词来组织句子成分,表达时、体、态和语气等。

由于缺乏形态变化,汉语形式化主要表现为虚词或语法标记词(grammatical markers)③的使用。因此,本文把虚词在翻译文本中的比例视为汉语翻译小说形式化的主要指标。如果我们把虚词的使用频率在总词频中所占的比例称为"形合度"(hypotactic level),那么,比较"形合度"的大小就可以得知语料库的句法显化程度:

表 11.2 各语料库的形合度比较

	CCTFC	LCMC(N)	LCMC	PDC2000_1④
词汇密度	53.48%	59.13%	63.33%	68.63%
形合度	39.54%	34.25%	29.64%	22.92%

---

① 由英国兰开斯特大学语言学系的 Tony McEnery 和肖忠华于 2003 年建成。LCMC 是一个严格按 Freiburg-LOB 英国英语语料库模式建设的现代汉语书面语平衡语料库。LCMC 的容量为 100 万词,包含 15 个文体类型,收入 500 篇字数约 2 000 词的样本,语料为 1989 到 1993 年中国内地的出版物,以 1991 年语料为主。其中包括五种类型的小说(McEnery and Xiao 2009)。

② 虽受欧洲语言影响,汉语也出现了类似于英语的"词缀"和复数形式,但数量非常有限,形式远没有英语那样复杂。

③ 汉语实词与虚词和各种词类的划分存在不同的划分标准和类型。如《马氏文通》把汉语分为实词和虚词两大类,"凡字有事理可解者,曰实字;无解而惟以助实字之情态者,曰虚字。"虚字分"介字、连字、助字、叹字"四类。黎锦熙《新著国文语法》将词分五大类:实体词、述说词、区别词、关系词、情态词。其中虚词是指关系词和情态词。王力《中国现代语法》认为词分为理解成分和语法成分,理解成分即实词包括:名词、数词、形容词、动词;语法成分包括半实词(副词)、半虚词(代词、系词)和虚词(联结词和语气词)。吕叔湘和朱德熙《语法修辞讲话》则把名词、动词和形容词归为实词,把代词、副词、连词、语气词、象声词等归为虚词。可见,争议最大的词类是副词和代词,究竟这两类词属于虚词还是实词,是汉语语言学家争论不休的问题之一。本章无意讨论汉语词类划分的问题,但由于在研究翻译文本与非翻译文本差异性时,不可避免地涉及汉语的词类问题,因此本章综合各家意见,主要采取了王力、吕叔湘与朱德熙先生看法,将名词、动词、形容词三类词归为实词;而将副词、代词、介词、连词、助词和叹词等归为虚词。

④ 兰开斯特大学开发的 2000 年《人民日报》语料库,本章所用为 2000 年 1 月语料。

（续表）

	CCTFC	LCMC(N)	LCMC	PDC2000_1
其他	6.83%	6.51%	6.85%	8.33%

表11.2对“形合度”进行了统计。为能更直观地反映各语料库的面貌，我们还统计了各语料库的词汇密度（即实词频率比例），并引入了新闻语料（PDC2000_1）供参考。表11.2中，LCMC是一个包含15种体裁，容量为百万词的现代汉语平衡语料库。它在一定程度上可代表现代汉语书面语的总体特征。我们且将LCMC的形合度视为现代汉语的“标准形合度”。从上表可见，小说文体LCMC(N)的形合度高于标准值，而新闻语料的形合度低于标准值。翻译小说的形合度为各语料库中最高，比标准值高出近10%，比非翻译小说高5%。这意味着，一部30万字左右中等长度的翻译小说要比同样长度的普通汉语著作多使用3万字的虚词。由此可见，翻译小说的语法标记显化程度高于汉语非翻译小说和一般汉语文体。

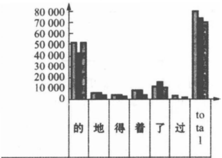
“形合度”从总量上体现了汉语翻译小说中语法标记显化这一基本特征。那么，语法标记显化是如何在翻译小说中具体实现的呢？显化究竟给翻译小说带来了怎样的实际效果？显化的来源是什么——是语言系统差异的必然结果，是译者在某种隐藏的规范指引下做出的共同选择，还是翻译作为一种特殊的认知活动所必然具备的共性？以下分别从助词、介词和连词的显化来描述翻译小说语法标记显化的实际表现，并试图寻求合理的解释。

11.3 翻译小说中语法标记的显化

11.3.1 助词的显化

我们统计了各语料库中主要的六个助词：结构助词“的、地、得”和动态助词“着、了、过”的频率。从图11.1中，我们可以看到这六个助词的绝对频率和语料库之间的相对差异：

从总频率来看，汉语小说助词使用频率高于总体，翻译小说高于非翻译小说和汉语总体。结构助词“的”的绝对值和相对差异最显著。“的”是现代汉语中词频最高的词，除作为结构助词引导定语和构成“的”字短语外，它还可以做语气助词表示强调。翻译小说中“的”的频率显著高于非翻译小说。从与汉语总体的比较来看，汉语原创小说倾向于较少使用“的”，而翻译小说却正好相反。这说明翻译小说中限定修饰成分较多。我们认为，在“的”的使用上，翻译小说体现出很强“显化”的趋势。



	的	地	得	着	了	过	to ta l
CCTFC	50	4,	3,	6,	11	2,	78
LCMC (N)	41	4,	2,	6,	14	1,	71
LCMC	50	3,	1,	3,	10	71	69

图 11.1 各语料库主要助词频率比较

（每百万词）

结构助词"地、得"的绝对频率较低，且翻译与非翻译小说之间相差不大。翻译和非翻译小说的动词频率无明显差别(胡显耀 2008：160)，与动词紧密相关的状语和补语也没有显著差异。

动态助词"着、了、过"中，"了"的绝对频率是最高的，这是由于在汉语中"了"的时体覆盖面更广，它既可表达行为的结束也可表示行为持续和状态持续，而"着"只能表示动作和状态的持续，"过"只能表示动作的完成和结束。从图 11.1 可见，由于小说以叙事、对话和描写为主，因而表达时体的动态助词使用频率高于汉语总体。从"着、了、过"的总体频率来看，翻译(20 709)反而低于非翻译(22 726)。换言之，翻译小说的时体语法标记没有非翻译小说显著。另外，值得注意的是，非翻译小说中"了"的频率较明显高于翻译小说。由于翻译小说的原语主要是英语等时体特征更明确(通过形态变化等)的语言，而汉语的时体特征较为模糊，三个助词都有不同程度的兼类现象(如上面的"了")。可见，在翻译过程中译者有意无意地淡化了原文的时体特征，使译文的时态不那么清楚。这是翻译文本的简化特征之一。

综上所述，在助词的使用上较为清晰的两个趋势是：翻译小说定语标记明显增多，而时体标记减少。我们把这种语法标记特征分别叫做翻译的"显化"与"简化"，其目的可能是降低翻译小说的难度，提高可接受性。

11.3.2 介词的显化

汉语介词是一个数量有限的封闭型词类。在语义层面上，介词的主要功能是"引出跟动词所指动作行为或形容性质有关的对象、施事、受事、时间、处所、方向、

方式、原因、目的、根据和范围等”(周小兵 1997)。在语法层面上,介词通常后跟词组构成介词短语主要充当状语、补语、定语和全句修饰语等句子成分(蔺璜 1997)。汉语的介词主要来源于古代汉语介词沿用、动词的虚化和外来语的影响。

贺阳(2004)曾通过语料统计指出,“关于、对、对于和在 N 的 V 下”等为“五四”以后新兴的介词和介词结构;“在、当”两个介词在“五四”以来使用频率明显升高;介词“当”的连词化等。如果说通过继承古代汉语的介词和动词的虚化是汉语自身的发展规律的话,贺阳(同上)提到的“现代汉语介词中的欧化现象”则是汉语与外来语接触的结果。由于语言系统的根本差异,汉语不会自动“欧化”,而是需要通过翻译来实现。因此,我们认为翻译小说中的虚词(包括介词)必然与汉语的传统语法有所区别。为了寻找这些差异性特征,我们对以下九个类别的 13 个最主要的介词在各语料库中的频率进行了统计和比较,见表 11.3:

表 11.3 各语料库主要介词词频比较(每百万词)

类别	介词	CCTFC	LCMC(N)	LCMC
施事	把	4013	2222	1985
受事	被	1217	1489	1243
与事	跟	918	576	237
	与	446	960	1429
工具	用	1115	746	984
方法	以	590	448	1353
处所	在	12956	9039	9881
	自(从)	202	218	225
对象	对	2978	2235	3149
	对于	210	128	295
范围	关于	115	77	255
目的	为(了)	1247	951	2206
原因	因(为)	1089	623	693
合计		27097	19711	23934

上表可分析如下:

第一,从统计的所有介词的总频率来看,翻译小说的介词使用频率明显高于非翻译小说,并高于汉语总体,而汉语原创小说似乎有少用介词的倾向。

第二,从个别介词的绝对频率来看,“在”的绝对频率在现代汉语中最高,“对”、“为(了)”、“与”分别次之。

第三，翻译小说中处所介词“在”和施事标记介词“把”的频率显著高于汉语小说和总体。而受事标记介词“被”和与事标记介词“与”的词频却低于非翻译小说。我们用图 11.2 来比较翻译和非翻译小说介词的频率。需要说明的是，以下柱状图的长短不反映各介词的绝对频率，而只是两个语料库之间的相对频率比较。

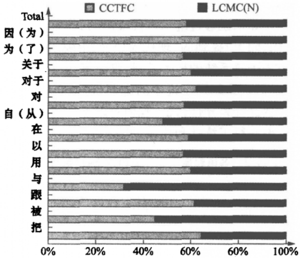


图 11.2 翻译和非翻译小说介词频率对比

从图 11.2 中可以直观地看到，翻译和非翻译小说中介词使用的差异：翻译小说明显更频繁地使用介词，介词总频率比非翻译小说高近 20%，多数介词的频率都高于非翻译小说，但有两个例外：一是被动标记“被”，二是与事标记“与”，且后者差距十分明显。关于“被”，我们将另文讨论。这里只考虑“与”。我们发现，同样是与事标记“跟”的频率却正好和“与”相反，它明显高于非翻译小说中的频率。我们研究了三个语料库中分别提取的索引句，发现这种现象可能跟“与”与“跟”的性质和语法功能有关：

首先，两者都可做与事标记介词，引入行为的陪同对象，一般情况下可以互换。如：

(1) 美国副总统到拉美修补与/跟拉美国家的关系。(A 新闻报道)[①]

(2) 此人的相貌与/跟模拟像特别像，作案手法又相近。(L 侦探小说)

① 例句(1)～(6)出自 LCMC 语料库，例句后括号中为语料类别。

其次，两者的区别在于："与"更多用于比较正式的、书面语的表达（如例(3)(4)），而"跟"更常见于非正式的、口语的表达（例(5)(6)）。"与"更多地用作连词（连接词语和短语），而"跟"基本上只用作介词。LCMC中"与"的频率高于"跟"，很可能是因为LCMC主要是书面语语料库。这一点可从汉语原创小说中"跟"的词频高于"与"得到证明。注意下列例句：

(3) 作为三十年代的女性，成家和思想感情的发展，不能与当时的潮流脱开。（P爱情小说）

(4) 高渐离奏雅乐而行刺的行为，更与燕国太子的事业无关。（G传记和杂文）

(5) 这样吧，我们还有点白切肉，这便跟客官们送来，再做碗菠菜场。（N武侠小说）

(6) 世雄，跟你伯父一屋睡。（N武侠小说）

翻译小说中"跟"的词频高于非翻译小说，体现了和其他介词同样的趋势。但"与"的频率却低于非翻译小说，我们对CCTFC中索引句的分析显示，"与"主要用于以第三人称视角描述事件（如(7)(8)(9)），而"跟"主要出现在人物的对话和自白式的叙述中（如(12)(13)(14)）。同时，"与"和"跟"的混用增多（如(10)(15)）。由于翻译小说"跟"的频率是"与"的两倍，而原创小说正好相反，所以可以理解为翻译小说中按照汉语小说的习惯可以用"与"的地方更多地使用了"跟"。

(7) 如今她出门到黑夜中去了，可他仍然与她同在。（《儿子与情人》）①

(8) 她仍与大学生来往，尤其与读物理化学的俄国青年交好。（《父与子》）

(9) 仿佛他与大山之间存在着某种神秘的联系。（《牛虻》）

(10) 她只与他谈过一两句话，而且当时他的身份是学校督察员。（《恋爱中的女人》）

(11) 珍妮呢，无论雷斯脱有病无病。当然乐意他跟自己一起的。（《珍妮姑娘》）

(12) 你再不要那样跟我说话了，好吗？（《海明威短篇小说选》）

(13) 我实在没办法，我是天生喜欢跟女人在一起。（《马丁伊登》）

---

① 例句(7)～15)出自CCTFC语料库，为节省篇幅，例句后只列出语料所出自作品的标题。作者、译者和出版信息参见胡显耀(2008:239-242)。

(14) 那个肌肉发达的男人站起来，跟她握了握手。(《本性》)

(15) "警察"臂章是信任他们可以跟德国宪兵来往的标志。(《青年近卫军》)

由此可见，"与"的减少和"跟"的增多在客观上起到了一种口语化和简化的效果。我们认为，这种做法很可能是一种降低文本的难度，提高翻译文本可接受性的做法——通俗地说，就是使译本显得更亲近一些。这是一种为使译语读者接受译本的考虑所做出的简化翻译策略。而介词的总体频率更高，也符合翻译小说的形合程度高于非翻译小说和汉语总趋势，这是语法标记显化的体现。

11.3.3 连词的显化

连词是一种具有多层连接功能的虚词，既可以连接词和短语，也可以连接小句和句子，还可以连接句子和句组(张斌、张谊生 2000:140-195)。汉语连词可以按各种标准进行分类，但比较通行的方法是，按意义将连词分为联合连词和偏正连词两大类，联合连词下又分并列、连贯、递进、选择关系的连词等小类；偏正连词又分因果、转折、条件、假设、让步、目的关系的连词等小类(同上:140-145)。按连接的语言形式，连词可分为连接词语和短语和连接分句和句子的连词；连接分句和句子的又分只能连接分句和既可以连接分句又可以连接句子，还可以连接段落的两小类(吕叔湘 1999)。由于连词的类型和数量较多，本文仅考察以下十类连词中最主要的连词在翻译和非翻译语料库中的表现(见表 11.4)。

表 11.4 各语料库主要连词频率比较(每百万词)

类　别	No.	连　词	CCTFC	LCMC(N)	LCMC
并列	1		112	102	152
	2		123	107	58
	3		68	38	27
连贯	4		12	43	46
	5	接着	365	192	95
	6	然后	513	431	291
递进	7	不但	39	81	93
	8	而且	666	303	394
	9	不仅	105	77	248
选择	10		27	4	4
	11		322	264	143
	12		22	9	8

（续表）

类 别	No.	连 词	CCTFC	LCMC(N)	LCMC
因果	13		1466	555	571
	14		271	171	461
	15		539	380	354
转折	16		460	533	382
	17		2433	1877	1788
	18		895	525	220
	19		297	256	217
条件	20		457	358	360
	21	只要	297	213	195
	22	除非	50	38	16
	23	不管	208	149	87
	24	无论	138	94	89
假设	25	如果	725	380	600
	26	假如	55	47	38
让步	27	即使	165	85	95
目的	28	以便	33	13	45
	29	以免	8	9	32
合计			10871	7332	7107

表11.4可分析如下：

第一，主要连词的总频率：小说文体中连词频率略高于汉语总体，但总体来说，汉语小说和非小说连词使用的频率差别不大，而翻译小说连词的频率显著高于非翻译小说和汉语总体。

第二，主要连词的绝对频率：各语料库连词频率总体趋势大致相同——绝对频率最高的联合连词是：而且、然后、是……还是；绝对频率最高的偏正连词是：但/但是、因为、如果。联合连词的词频总数低于偏正连词，这可能是因为汉语并列、承接、递进和选择关系的复句倾向于不使用连词（见图11.3和图11.4）。

第三，主要连词的相对频率：翻译小说中绝大多数的连词频率都高于非翻译汉语，例外只有："首先、不但、不仅、以便、以免"等五个连词。翻译小说的频率明显高于非翻译小说的连词是："因为、但(是)、接着和而且"，这几个连词同时又是各语料库中高频连词。由于"不但、不仅"两个连词通常与"而且"合用，故可以认为翻译小说中单用"而且"更为频繁。汉语小说和汉语总体在连词使用上看不出明显差异：一般来说，小说中的连词频率高于总体，但在递进关系连词"不但、不仅、而且"三词的表现上，小说的频率低于总体。

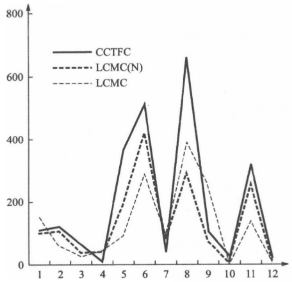


图 11.3 各语料库联合关系连词频率比较

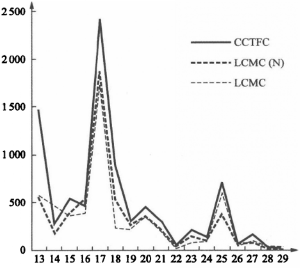


图 11.4 各语料库偏正关系连词频率比较

11.4 小结

通过上述分析，我们认为："语法标记显化"是翻译小说区别于非翻译小说和其

他汉语原创文本的显著特征之一。主要的语法标记词:助词、介词和连词的使用特征都分别证实了这一趋势。同时,我们还发现,翻译小说定语和语气标记词("的")明显增多,而时体标记词("着"、"了"、"过")减少。此外,介词"与"的减少和"跟"的增多在客观上起到了一种口语化或简化的效果。翻译小说中的连词在总体频率上明显高于非翻译小说,其中因果、转折、承接和递进关系连词增加较显著。

翻译小说语法标记显化的原因固然可以解释为是源语(尤其是高形式化的源语)的影响所致,即翻译汉语中普遍存在对高形式化源语的"仿译"。但正如第 11.1 节所言,外译汉过程中句法显化是非强制性的,为什么现代汉语翻译小说的译者总体上选择了"显化"呢? 译者如何在原语和译语两种截然不同语言规范中做出选择是语言系统差异无法解释的。

因此,另一种解释是:"显化"是译者在社会因素控制下所做出的选择,或者说,语法标记显化是"翻译规范"控制译文生产的结果。形合的客观结果是使译文趋向于更贴近原文而偏离译语。对译语文化来说就是趋向于翻译文本的陌生化,希望翻译能向译语文化提供"异质"的新材料以丰富和更新译语文化系统并促进其发展。当然,如果某一时期译语社会对翻译文本的期待倾向于"隐化",那么这一时期的翻译文本就应该出现语法标记隐化的总体特征。不过,这一假设需要对大量翻译文本的历时性考察才能得以证实。

我们还可以从翻译行为的认知特性出发,做出第三种解释:语法标记显化可能是汉语翻译文本的普遍特征或"翻译共性"之一。翻译的理解认知过程,翻译所隐含的解释性,以及提高译文可接受性这一基本倾向都可能造成译者提高译文形式上明确程度,减少歧义,简化语义等等。第三种解释更贴近于"显化"作为翻译共性这一假设,不过,目前语料库翻译研究的局限性在于——它可以提供实证的证据说明翻译文本的总体趋势,但它本身却无法对这些趋势加以解释。因此,本文所描述的显化现象还有待翻译学界更多的解释性研究。

第 12 章

英译汉文本中人称代词的显化特征

本章借助双语平行语料库对英汉翻译中人称代词主语在文学和非文学两种文体类型中数量、频次和转换类型三方面考察发现：文学与非文学英译汉时，人称代词主语数量和频次均呈减少趋势；从转换类型看，人称代词主语转换以对应关系为主，语际显化和隐化均不明显；相对非翻译汉语文本，汉语翻译文本类比显化突出。研究表明：英译汉过程中，人称代词主语语际转换表现出源语迁移现象；语内类比显化突出，根本原因在于英、汉语对形式手段依赖程度的差异，以及语言相对社会地位和认知规律等因素的作用。

12.1 引言

语料库翻译研究（corpus-based translation studies）途径经过十多年的理论阐述与实证研究已经发展成为一种新研究范式，研究内容涉及翻译语言共性特征、译者文体、翻译转换过程等，其中以对翻译共性（translation universals）的研究最为突出，翻译共性中又以对显化（explicitation）的讨论较为集中。由于英、汉两种语言对人称代词手段在构句层面的依赖程度不同，英汉翻译中人称代词的转换表现出一定的特点。限于篇幅，本章借助汉英平行语料库，仅考察人称代词主语（且仅限于人称代词主格形式，不包括物称代词、反身代词和领属代词）在英汉翻译中的转换，以此入手探究英汉翻译中的显化现象，并尝试从语言形式手段差异的角度加以解释。

12.2 英汉翻译中的人称代词

Quirk et al（1985：347）指出，英语第一、二人称代词有具体所指时通常用来直接指称会话场景中的说话者和称谓对象，第三人称代词则用于指称能根据语言外因素推断出的人或事物，更多的是用于预指或回指。吕叔湘（1999：8）指出："汉语里可以不用人称代词的时候就不用；即使因此而显得句子结构不完整，也不搞形式主义"。可见，人称代词"隐去"（implicating）是汉语的一个特点。相比较而言，英

语对人称代词的依赖程度要高于汉语。已有文献对此做过探讨，我们又从语料库数据分析进一步加以论证。

赵世开(1996,1999)对英汉和汉英翻译中人称代词使用频次考察发现：无论英译汉还是汉译英，1)英语第三人称代词比汉语用得多；2)英语第一人称单数代词比汉语的使用率低；3)第二人称代词的使用汉语比英语多，据此他以说话人(S)和听话人(H)为语用参数将英语和汉语区分为 X 型和 Y 型[①]。根据赵世开(1999)的统计：就人称代词数量看，从英语原文到汉语译文，第一人称代词减少，第二人称代词增加，第三人称代词减少最突出；从汉语原文到英语译文，第一人称代词增加，第二人称代词增加不十分明显，第三人称代词增加最突出；从人称代词所占比例上看，从英语原文到汉语译文第一人称代词比例增加不明显，第二人称代词比例明显增加，第三人称代词减少；从汉语原文到英语译文，第一人称比例减少不明显，第二人称减少，第三人称增加突出。这一研究将汉英对比研究与翻译研究相结合，具有开创性，为后来就英汉、汉英翻译中人称代词转换类型研究，以及对人称代词增加或减少在译文中的作用研究奠定了基础。

根据刘礼进(1997)对一部英语小说及其汉译本第一章中第三人称代词频次的统计，汉语译文中的第三人称代词频次低于英语原文，刘礼进指出英语回指式人称代词的频次高于汉语，但汉语译文中回指式人称代词的频次也很高。这一研究未对翻译文本(translated text)和译语非翻译文本(TL non-translated text)中第三人称回指的情况加以对比分析，因此还不能确认英语回指式人称代词的频次高于汉语。

赵宏、邵志洪(2002)选择汉、英各两个小说文本及其译本各约一千字，对其中第三人称前指照应进行统计发现，英语语篇中第三人称代词的使用频率高于汉语，而且这一倾向在汉译英文本中表现更为突出；汉译英文本中人称代词的添加十分突出，英译汉文本中的省略不大明显，相比而言，保留原文中人称代词的做法反而十分明显。

王琼琼、王克非(2004)从第三人称代词作为照应手段的视角出发，发现学生英汉翻译习作中会出现汉语第三人称代词的冗余现象，但学生以汉语母语写作却不会出现第三人称代词的冗余现象；汉英翻译中，英语译文中对第三人称代词的使用也大大高于汉语原作中单数第三人称代词的频次，即发生了人称代词的添加现象。

以上研究对英汉或汉英翻译过程中人称代词使用进行了关注，但尚未对人称代词在翻译中的转换类型与同类非翻译文本中的情况加以对比，未就人称代词在

---

① 第三人称出现比例高的是 X 型，第一、二人称出现比例高的是 Y 型；说话人单数人称代词使用频次低，复数使用频次高的为 X 型。相反，单数人称代词使用频次高、复数使用频次低的为 X 型(赵世开 1999：20)。

翻译转换过程中的对应(correspondence)、明示(explicating)、添加(adding)或隐去(implicating)等进行深入探讨,而且用于考察人称代词使用的语料大都限于文学文本,并且以人工统计为主的语料规模较小。我们在前人研究的基础上,从人称代词主语在英汉翻译中的表现为考察对象,以"汉英对应语料库"提供的英语原文及其对应汉语译文为基础,并且以汉语非翻译文本作为类比语料,探索英汉翻中的显化。

12.3 英汉翻译中人称代词主语转换的语料库考察

12.3.1 翻译中的显化

作为翻译共性的一个方面,本章对显化现象的探讨主要集中在两个方面:第一,是指翻译过程中,在译文中添加或明示原文中隐含(implicit)语言成分的过程,目的在于更清楚地传达原文中的语法和非语法信息,相反的过程则被称为隐化(implicitation),即根据译语语言规范隐去原文中语言成分的过程。另一类显化是指翻译文本相对于译语中非翻译文本所表现出的显性程度(explicitness)提高。本章将前一种显化视为语际对比基础上的显化,简称语际显化(inter-language explicitation);而将仿译式的显化视为翻译语言相对译语中非翻译语言所表现出的语内类比显化,简称类比显化(comparable explicitation),此两者都应作为显化研究的对象。Chesterman(2004a,2004b)将其分别归入源语型共性(S-universals)和目标语型共性(T-universals)中(吴昂、黄立波 2006;黄立波、王克非 2006 等)。同时需要指出的是,翻译转换中的显与隐通常相伴而存①,本章将语际隐化也作为关注的一个方面,限于篇幅,类比研究中主要以对类比显化的探讨为主。

12.3.2 人称代词主语在翻译中的转换类型

英汉翻译中,人称代词主语的转换类型可分为对应、明示、添加和隐去四种,具体如下②:

第一,对应即原文中的人称代词主语在译文中同样以人称代词主语将其译出。如:

(1) Using these sources as my guide, I retraced what I believe was the path he took from Bellingham to Madison County in August of 1965. Driving

① 语际显化与语际隐化两者并不存在对称关系(Klaudy & Károly 2005)。

② 源语与译语例句中的人称代词主语均以黑体标出。

toward Madison County at the end of my travels, I felt I had, in many ways, become Robert Kincaid.

以这些材料为线索，我沿着我认为是金凯 1965 年 8 月从贝灵汉到麦迪逊县的路线作了一次旅行，在行程终了时，我觉得自己在很多方面变成了罗伯特·金凯。(Robert James Waller, *The Bridge of Madison County*，梅嘉译)

(2) To spare himself, he agreed to disarm of all weapons of mass destruction.

为了保全自己，他同意解除全部大规模毁灭性武器。(George W. Bush, *State of the Union Message*)

第二，明示是指将原文中的逻辑主语在译文中以人称代词主语的形式译出。如：

(3) Edward found it something of an effort to comply, for his mind kept wandering—trying to remember what the service was that he had done Goodson.

听着妻子的话，爱德华有点儿跟不上，因为他总是心猿意马——他在使劲想到底给古德森帮过什么忙。(Mark Twain *The Man that Corrupted Hadleyburg*，常健译)

(4) The claimant Robison was a conscientious objector who had performed alternative service, as required by the draft law.

申请人鲁宾逊是个因道德或宗教原因而拒服兵役者，他提供了《征兵法》所要求的其他可替代的服务。(R. W. Findley & D. A. Farber, *Environmental Law in a Nutshell*)

第三，添加即原文中没有出现主语，但在译文中却根据上下文添加了人称代词作为主语。如：

(5) "It was wicked, that is what it was—just blasphemous presumption, and no more becoming to a meek and humble professor of—"

"这叫瞎折腾，就是这么回事——敬酒不吃吃罚酒，你就别再装老实人、装规矩人啦——"(Mark Twain, *The Man that Corrupted Hadleyburg*，常健译)

(6) The agency also adopted a summary judgment procedure, under which a manufacturer facing disapproval of a drug would be denied a hearing unless it could demonstrate in advance that it could present a "genuine and substantial issue of fact" under the new rules.

该机关还采取了即决审判程序，依此程序，面对产品遭否决的制造商将被剥夺听证机会，除非他事前可以证明，根据新的规则，他可提出“真正和实质的事实问题”。(R. W. Findley & D. A. Farber, *Environmental Law in a Nutshell*)

第四，隐去是指原文中的人称代词主语，在译文中不予译出，可根据交际语境或上下文加以判断。如：

(7) The words had been there for years, and she had never said them.

这句话已存了多年，但是从来没有说出来过。(Robert James Waller, *The Bridge of Madison County*，梅嘉译)

(8) "What do you have to eat?

"有什么吃的东西？(Ernest Hemmingway, *The Old Man and the Sea*，吴劳译)

这里需要指出的是，有时候明示与添加之间的界限较为模糊，但因为两者均为显化的表现形式，因而并不影响数据的统计和研究结果。

12.3.3 英汉翻译中的人称代词主语转换

以下从文学和非文学文本两种文体类型来考察人称代词主语在英汉翻译中的数量、频次和转换类型。所选英语原文语料约 18 万词，其汉语学文本约 8 万词，非文学文本约 10 万词(见表 12.1)。

表 12.1 英汉翻译中人称代词主语考察所选语料

文体类型	文本名称	原文字/词数	译文字/词数
文学文本	*The Bridge of Madison County*	36 620	63 221
	The Old Man and the Sea	26 920	36 705
	The Man that Corrupted Hadleyburg	18 198	34 639
总计		81 738	134 565
非文学文本	*State of the Union Message*	5 570	2 300
	Administrative Law	96 971	197 690
总计		102 541	199 990

文学文本包括 *The Bridge of Madison County*(《廊桥遗梦》，Robert James Waller 著，梅嘉译)、*The Old Man and the Sea*(《老人与海》，Ernest Hemmingway 著，吴劳译)和 *The Man that Corrupted Hadleyburg*(《败坏了赫德来堡的人》，

Mark Twain 著，常健译)；非文学文本包括 *State of the Union Message*(《国情咨文》，George W. Bush)和 *Administrative Law and Process in a Nutshell*(《行政法概要》，Ernest Gellhorn & Ronald M. Levin)①。

同时以汉语原创文本作为参照库，所选汉译英汉语原文语料约17万字，其汉语学文本约9万，非文学文本约8万字(见表二)，基本保证了类比语料的平衡。

表 12.2 汉语非翻译文本参照库所选语料

文体类型	文本名称	原文字/词数
文学	《围城》1～3章	65 386
	《骆驼祥子》1～3章	16 770
	《春蚕》	12 693
总计		94 849
非文学	《政府白皮书2》	71 956
	《在中国共产党全国代表会议上的讲话》	3 726
总计		75 682

文学文本包括《围城(1～3章)》(钱钟书著，Jeanne Kelly & Nathan K. Mao 译)、《骆驼祥子(1～3章)》(老舍著，施晓菁译)和《春蚕》(茅盾著，Sidney Shapiro 译)；非文学文本包括《政府白皮书2》和邓小平《在中国共产党全国代表会议上的讲话》②。

12.3.3.1 英语原文与其对应汉语译文中人称代词主语比较

以下首先从数量和频次两方面考察英语原文与对应汉语译文中人称代词主语的使用情况。借助"汉英对应语料库"的自动检索工具和词频统计功能，同时对非主语的人称代词形式加以排除，英语原文和对应汉语译文中人称代词主语的数量与频次统计如下(见表12.3)：

表 12.3 英语原文与其汉语译文中的人称代词主语

文体类型	文学文本		非文学文本	
文　本	英语原文	汉语译文	英语原文	汉语译文
大小(字/词)	81 738	134 565	102 541	199 990
人称代词主语	5 251	3 905	352	242
频次(‰)	64.24	29.02	3.43	1.21

① 非文学文本统一不注明译者。

② 限于篇幅，本章节选《围城》和《骆驼祥子》两个文本各三章内容作为语料；《政府白皮书2》共包含六个文本：1)《中国海洋事业的发展》；2)《中国的劳动和社会保障状况》；3)《中国的农村扶贫开发》；4)《西藏文化的发展白皮书》；5)《中国的航天白皮书》；6)《中国知识产权保护状况》。

根据表 12.3，英语文学文本原文约 8 万词，人称代词主语出现约 5 251 次，频次为约每千词 64 次，对应汉语译文约 13 万字，人称代词主语约 3 905 个，频次为约每千字 29 个；非文学文本英语原文约 10 万词，人称代词主语的频次为 3.43‰，而对应汉语译文为 1.21‰。

从数量与频次上看，英汉翻译中人称代词主语的数量与频次均有所降低（如图 12.1 所示）。

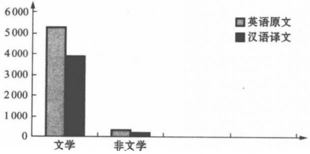


图 12.1 英语原文与汉语译文中的人称代词主语

英汉转换中人称代词主语在数量与频次均表现出降低的趋势，文学翻译中降低幅度为 35.22‰，非文学翻译中降低 2.22‰。由此可初步说明，若将人称代词主语视为语言形式化程度的一个标志，那么汉语对人称代词主语的依赖程度通常要低于英语，这一点还有待于下文进一步论证。

12.3.3.2 英译汉过程中的人称代词主语转换

以下对英译汉时人称代词主语的具体转换类型加以考察，主要考察英语人称代词主语在翻译到汉语过程中所表现出的对应、明示和添加现象（见表 12.4 和表 12.5）。

表 12.4 英译汉文学文本人称代词主语转换类型

文本名称	总句数	人称对应		语际显化			
				明示人称		添加人称	
		句数	比例（%）	句数	比例（%）	句数	比例（%）
《老人与海》	1 323	1 283	96.98	28	2.12	12	0.91
《廊桥遗梦》	18 34	1 773	96.67	49	2.67	12	0.65
《败坏了赫德莱堡的人》	748	627	83.82	65	8.69	56	7.49
总计	3 905	3 683	94.31	142	3.64	80	2.05

根据表 12.4，英译汉文学文本中人称代词主语转换类型主要以对应关系为主，占到总考察句数的 94.31%，远远高于明示与添加之和 5.69%。

表 12.5 英译汉非文学文本人称代词主语转换类型

文本名称	总句数	人称对应		语际显化			
				明示人称		添加人称	
		句数	比例(%)	句数	比例(%)	句数	比例(%)
《国情咨文》	135	132	97.78	3	2.22	0	0
《行政法概要》	107	67	62.62	19	17.76	21	19.63
总计	242	199	82.23	22	9.09	21	8.68

非文学文本中，人称代词主语转换情况与文学文本相同，对应关系占 82.23%，大大过于明示和显化之和 17.77%。对英汉翻译中人称代词主语的转换类型按照对应与显化(包括名示与添加)加以归纳(如图 12.2 所示)：

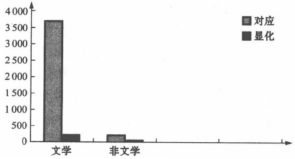


图 12.2 英译汉文学与非文学翻译中的人称代词主语转换

无论文学还是非文学文本，就所考察的英汉翻译中人称代词主语转换来看，对应关系表现突出，语际显化不明显。根据图 12.2 还可看出，英语文学文本中对人称代词主语的使用比较丰富，而对应汉语译文中却大大减少，从这一点也体现出英、汉语对人称代词主语的依赖程度差异，结合以上汉英翻译中关于人称代词主语的讨论，从一个侧面也说明了汉语在构句层面上意合的特点。如果可以将人称代词主语作为语言形式化程度的一个参数，那么与英语相比汉语的形式化程度相对要低一些。英汉翻译中人称代词主语的语际隐化表现如下(见表 12.6 和表 12.7)：

表 12.6 英汉文学翻译中人称代词主语的语际隐化

类型		英语显性	英汉对应		语际隐化	
人称代词主语	Ⅰ人称	1570	1141	72.68%	429	27.32%
	Ⅱ人称	423	347	82.03%	76	17.97%
	Ⅲ人称	3258	2195	67.37%	1063	48.43%
总计		5251	3683	70.14%	1568	29.86%

表 12.7 英汉非文学翻译中人称代词主语的语际隐化

类型		英语显性	英汉对应		语际隐化	
人称代词主语	Ⅰ人称	118	116	98.31%	2	1.69%
	Ⅱ人称	10	6	60%	4	40%
	Ⅲ人称	224	77	34.38%	147	65.63%
总计		352	199	56.53%	153	43.47%

根据表 12.6 和表 12.7，英语原文中的显性人称代词主语在翻译转换中总体上都以对应关系为主，文学文本中为 70.14%，非文学文本中为 56.53%，语际隐化分别占 29.86%和 43.47%。需要指出的是，非文学文本中第三人称代词主语在英汉转换中表现出语际隐化的趋势，主要原因在于第三人称回指在英语中比较广泛。可以发现：英汉翻译中人称代词主语的转换，数量与频次上呈递减趋势，转换类型上则以对应关系为主，语际显化不明显，语际隐化总体不明显。

12.3.3.3 汉语翻译文本与非翻译文本中人称代词主语比较

上文分别对人称代词主语在文学与非文学两种文体类型中的数量、频次和转换类型进行了考察。初步表明：文学与非文学文本中都表现出人称代词主语的数量和频次的减少；就转换类型看，语际显化和语际隐化均不大明显。以下通过对汉语翻译文本和非翻译文本中人称代词主语使用的比较，来讨论类比显化现象。汉语翻译文本与汉语非翻译文本中人称代词主语的频次表现如下(见表 12.8)：

表 12.8 汉语翻译文本与非翻译文本中人称代词主语比较

文体类型	文学文本		非文学文本	
文　本	汉语翻译文本	汉语非翻译文本	汉语翻译文本	汉语非翻译文本
大小(字/词)	134 565	94 849	199 990	75 682
人称代词主语	3 905	1 608	242	29
频次(‰)	29.02	16.95	1.21	0.38

由表 12.8 可以看出，文学文本中，人称代词主语在翻译文本和非翻译文本中的频次分别为 29.02‰和 16.95‰，平均每千字多出约 12 个；非文学文本中，分别为 1.21‰和 0.38‰，平均每千字约多一个。由此可以看出，无论文学还是非翻译文本，与汉语非翻译文本相比，汉语翻译文本中人称代词主语的使用频次要高。结合以上讨论可以做出推论：英汉翻译中，翻译文本由于受原语文本的影响，表现出人称代词主语迁移，导致其频次的增加。人称代词的回指和预指的作用是明示语篇中人际关系的主要手段之一，因此人称代词主语的增加说明汉语翻译文本呈现类比显化的趋势。

12.4 对翻译中显化的认识

通常认为，翻译转换中所表现出来的一些特征完全是由于语言系统具体差异的结果，但基于语料库的实证研究（Øveras，1998；Olohan & Baker，2000 等）表明翻译中的显化在许多情况下确属于一种非强制性的策略，即并非译语文本的绝对要求。如前文所讨论的英汉翻译中人称代词主语在译文中添加或明示与否，均不是译文语言系统所强制要求的。但不同语对（language pair）间的翻译是否具有相同的共性呢？现有实证研究仅限于具体语对中就特定语言项的考察，难免停留在对局部共性的研究上。抛开具体语对的约束，可以从语言形式化程度的角度将这些局部共性统一起来。这里所谓"语言形式化"（linguistic formalization）不同于语言分析方法的"形式化"（formalization）。语言分析方法的"形式化"是指语言学，特别是生成语言学中"支配一种分析法的规则、原则、条件等能以精确和严格的方式界定"，即"能用逻辑或数学术语对一种语言分析做出形式解释并制定演算方法"（克里斯特尔 2000：146）。换句话而言，语言分析方法的"形式化"就是对语言分析方法的逻辑化和数学化。语言形式化是指某一语言表现某种意义关系所依靠的各种方式，包括音调、语序、屈折变化及功能词等的丰富程度，以及该语言对这些形式手段的依赖程度。吕叔湘（1999：7）提出："汉语语法的最大特点是没有严格意义的形态变化"，"没有严格意义上的形态变化，在语法分析上就容易引起问题"。吕叔湘、朱德熙（1979：10）指出，正是"由于汉语缺少发达的形态，许多语法现象就是渐变而不是顿变，在语法分析上就容易遇到各种'中间状态'"。这里所谓"发达的形态"是指西方语言中的各种形态变化形式（如性、数、格、时、体、语态、词类标记等），相对而言，汉语在这方面比较缺乏，使范畴与范畴之间并非泾渭分明，造成语言分析的困难。目前，对语言形式化程度这一问题的探讨较少，但不同语言在以各种形式手段表达意义及意义关系方面均表现出各自的典型特征，这一现象客观存在。而且从语言本质差异入手，对揭示翻译转换中各种现象的本质举足轻重。从语法角度讲，语言形式化程度是指一种语言中基本形式手段的多寡，和该语言依赖各种形式手段表达语义关系的程度，形式手段多并且对其依赖程度高的语言，我们就说其形式化程度高，反之则形式化程度低。

柯飞（2005：307）指出，翻译的中显化和隐化可由多种因素造成，具体包括语言因素、文本因素、译者因素和社会文化因素。其中前两类属于语言内因素，后两者是语言外的因素，具体包括人类认知规律和语言相对社会地位的差异两个方面。从人类的认知规律上讲，我们总是依靠已知的东西来认识未知的东西，这一点体现在翻译中就是显化现象。另一方面，语言使用中的经济原则又要求以最节约的形

式来表达各种意义和意义关系，在翻译中则体现为隐化。语言之间的相对社会地位，是指不同语言之间在不同社会时期由于政治、经济、国力等因素在相互交流所表现出的不同地位，就翻译而言具体可体现在翻译策略和翻译产品的“顺差”与“逆差”问题两方面。从弱势语言向强势语言的翻译（如汉译英、芬译俄等），语际显化的趋势相对比较突出；而由强势语言向弱势语言的翻译中，翻译文本的类比显化特征则比较明显。对于显化应当注意以下几个问题：第一，显化并不一定就意味着添加，而且并非所有的添加都与显化有关；第二，不能从本质上将语言划分为明晰或含蓄，因为不同语言会在不同层面上表现出明晰或含蓄的特点，显化或隐化存在于翻译转换过程中；第三，至少在句子层面，添加与省略在翻译中总是并存的，但显化或隐化与否还需要从不同的层面加以识别，如句子层面的显化并不意味着语篇层面的显化。

12.5 小结

通过数量、频次和转换类型三个方面的考察发现：英汉翻译中，人称代词主语语际以对应关系为主，抛开译者语言能力因素，表现出英语作为强势语言在译语文本中的迁移现象，造成译语中的冗余现象；与原创的汉语文本相比，汉语翻译文本表现出类比显化的趋势。倘若将对人称代词的依赖程度作为语言形式化程度的一个参数，那么由形式化程度高的语言翻译到形式化程度低的语言时，类比显化突出，即翻译译文在形式上比译语中的非翻译文本显性程度高；从文体类型上讲，文学与非文学文本翻译中的类比显化表现均比较高。这一研究结果进一步说明语言形式化程度是影响翻译转换中形式方面显化与否的重要因素。作为一种翻译共性，此种形式上的显化现象在其他语对中亦有所表现（如匈译俄、匈译英、挪译英、英译葡等），但还需要进一步的考察。

第13章

英译汉文学文本中人称代词的变异

本章通过汉语翻译文学与汉语原创文学语料的对比和分析，探讨汉语翻译文学文本中人称代词使用的特征。本章发现：1）汉语翻译文学中，各类人称代词的使用频率均高于原创文学；2）第三人称代词“他”的复现率明显提高；3）“他”在汉语文学翻译语料中的照应功能明显增强，出现了不同指“他”交替的不同于汉语语法常规的变异特征。我们认为：人称代词语法显化和变异是汉语翻译文学文本的显著特征之一。这些特征可能反映了英语等形态比较丰富的语言对翻译汉语的影响，体现了现代汉语翻译文学作品的陌生化操作规范。

13.1 引言

现代汉语中，代词除具有指别功能外还有照应功能，即与其他语言单位构成同指关系的功能。王力（1989）把代词归入“半虚词”，就是指它的照应功能。人称代词的照应功能可按代词与名词性成分的关系分为同指和异指；按先行词与代词的位置关系分为前向照应和后向照应，或称回指（anaphor）和前指（cataphor）（徐赳赳，2003）；按先行词与代词是否同在一个句子中分为句内照应和超句照应（董红源2002：6-8）。代词有三类：人称代词、指示代词和疑问代词；人称代词又分第一、二、三人称（“三身代词”），指示代词又分近指和远指。与英语及其他印欧语言相比，汉语代词的类型较少，使用频率较低（刘宓庆1998：445-476；范仲英1997：148-151）。汉语人称代词形式上没有主格宾格之分，代词所有格与英语的物主代词相比使用得更少。汉语常规的指代方式主要以“名词复现”和“零代词”为主，显性人称代词的使用频率一般较低。

据Francis和Kučera（1982：547）、Johansson和Hofland（1989：15，转引自Leech，2001）对BROWN和LOB的统计，“想象型”语料比“信息型”语料明显更频繁地使用人称代词。胡显耀（2008：143-147）对汉语语料词汇特征的考察也表明汉语小说比其他文体更倾向于使用更多的人称代词。这说明多使用代词是英、汉语文学语料的共同特征之一，原因是以小说为主的文学语料通常以对话和叙述为主，

而叙述通常以第三人称和第一人称为主。

在使用北外"英汉对应语料库"对翻译文学语料的人称代词的考察中，我们发现：人称代词在汉语翻译文学语料中的使用频率、指代功能、照应方式等具有明显不同于汉语非翻译(原创)语料的特征。总体而言，汉语翻译文学语料中人称代词的作用更类似于英语的语法功能词或虚词，换言之，汉语翻译文学语料比原创文学语料更倾向于指代显化(explicitation)，而不是遵循汉语常规，使用名词复现和零形代词的方式。此外，翻译汉语中人称代词的指代功能显著增强，在照应方式上出现了不同指的"他"同句和后向照应等一些汉语原创文本中少见的变异特征(deviation)。

13.2 人称代词的显化

根据语料库的词性赋码，我们分别统计了各种人称代词在汉语原创文学语料，与之对应的英语翻译文学语料，英语原创文学语料和与之对应的汉语翻译文学语料中的使用频次并标准化了各语料库每百万词中人称代词的数量。

由于英汉语人称代词系统的差异，我们对统计数据进行了修正[①]。为了全面地了解两种语言的代词使用特征，我们将非人称代词"它"和"it"也列入了考察和比较范围。表13.1统计了各语料库中，主要人称代词及其复数形式和所有格的频率。

表 13.1 汉英文学(翻译)语料库主要人称代词词频比较

每百万词	原创汉语	翻译汉语	原创英语	翻译英语
我(I,me)	8300	14267	15592	7999
你(你们,you)	7634	8811	8984	10952
他(he,him)	11726	17328	19149	15950
她(she,her)	5266	6467	8706	11642
它(it)	531	2058	14600	8035
我们(we)	1716	2787	2644	3007

① 由于英语和汉语人称代词的形态不尽相同：汉语无主格宾格之分；英语 you 的单复数同形；英语的第三人称复数有 they 和 them 两种形式，而汉语有"他们、她们和它们"三种形式；英语的人称代词还有反身代词形式，而汉语以人称代词加"自己"来表示，等等。为了使两种语言的人称代词具有可比性，我们对统计各语料库的数据进行了修正，即"我"的英语对应频率应该是"I"及其宾格"me"之合；英语的 you 对应汉语"你"和"你们"；they/them 之合对应"他们、她们和它们"等等。此外，由于汉语人称代词所有格是通过加"的"来实现的，因此我们从相应形式中扣除了人称代词加"的"的数量。

（续表）

每百万词	原创汉语	翻译汉语	原创英语	翻译英语
他们（they, them）	2 781	4 716	5 967	5 615
你的（your）	517	709	1 524	1 918
我的（my）	622	1 358	3 412	1 986
他的（his）	1 477	2 692	11 299	12 991
她的（her）	763	1 030	6 303	7 756
它的（its）	78	259	1 421	724
合计	41 411	62 482	99 601	88 573

对表 13.1 的数据进行分析，我们可以发现一些英汉语人称代词使用的系统差异，但更重要的是，汉语翻译文学与原创文学语料相比，人称代词使用中呈现的特殊性：

13.2.1　英汉语人称代词使用的系统差异

从表 13.1 中的主要人称代词总频率来看，原创英语文学语料中人称代词总频率（99 601）是原创汉语文学语料（41 411）的 2.4 倍。图 13.1 列举了三个语料库的主要人称代词的词频分布：最下面的曲线是原创汉语文学语料的人称代词频率，最

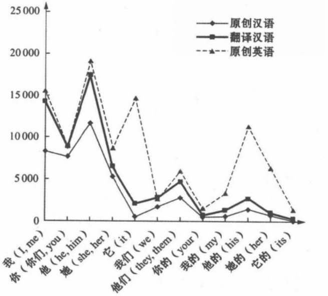


图 13.1　翻译汉语与原创英汉语人称代词频率比较

上面是原创英语文学语料人称代词频率的变化曲线。可见,原创汉语全部人称代词频均显著低于原创英语。

尽管已经有很多研究表明英语的人称代词使用频率高于汉语,但语料库的量化研究使我们能够更准确地发现汉语比英语少用人称代词的具体程度和在不同形式间的分布。除总词频外,较明显差别还有:

第一,英语"it"的词频是汉语"它"的27.5倍,这是由于it在英语中除充当代词外,还有作为先行词、虚义词和强调词等功能(连淑能,1993:81-82)。

第二,英语第三人称"he/him"及其所有格"his"的使用频率与汉语相比有明显差别。其中"he/him"为"他"的1.6倍;"his"为"他的"的7.6倍。这一方面说明汉语文学语料通常较少使用人称代词所有格,另一方面的原因可能是his在英语中可同时兼任形容词性和名词性所有格。

第三,原创英语文学语料第一人称代词(I及其宾格me)总使用频率明显高于汉语"我"。

第四,原创英语文学语料第三人称"he/him"的词频高于其他英语人称代词,是"she/her"的两倍;而汉语"他"虽略多与"她"但无明显差别。

13.2.2 汉语翻译文学与原创文学人称代词使用频率差异

图13.1中间实线为翻译汉语人称代词频率曲线。可以清楚地看到,除了由于汉语语法规则的限制,"它"和"他的"远低于英语的"it"和"his"的频率之外,汉语翻译文学语料中人称代词的分布规律与原创汉语相近,但翻译汉语全部人称代词的使用频率皆高于原创汉语。较显著的差异是:1)第一人称单数"我",翻译汉语是原创汉语的1.7倍;2)第三人称单数"他",翻译汉语是原创汉语的1.5倍。

不少论者在谈到代词翻译的时候提出,英语倾向于更频繁地使用代词,而汉语更频繁地使用原词复现和"零形回指"(即省略代词或名词)的方式,因而在翻译中应少用代词多用原词复现等方式(冯庆华1997)。但从对翻译文学语料中人称代词的实际使用的统计分析来看,汉语翻译文学中人称代词的实际使用频率显著高出原创文学,除it和所有格his/her外,人称代词的使用较接近于英语。我们把汉语翻译文学中人称代词的这一特征称为人称代词的"显化"。显化主要说明了翻译汉语中人称代词使用的频率显著提高,但频繁使用人称代词究竟会给翻译文本造成怎样的影响?以下我们详细考察人称代词"他"的使用及其效果。

13.3 人称代词的变异:以"他"为例

语言学界的人称代词回指研究表明:第三人称代词"他"在汉语中具有特殊作

用。罗选民(2001:221)认为:"(人称指称)在汉语中最常见的是使用第三人称代词单复数形式";王灿龙(2000:228)也提出:"在现代汉语系统中,人称代词'他'是使用频率最高的照应语之一('他'实际上也指'她'、'他们'、'她们')。"(以上转引自赵宏、邵志洪 2002;封宗颖、邵志洪 2004)。徐赳赳(2003:106-137)对"他"的格分布、语序列和延续性进行了统计研究。他认为照应语之间的语序列通常是名词最先出现接着以人称代词、零形式或名词复现的方式进行回指,而代词和零形式在前,名词后出现的前指情况在汉语中虽然存在但非常少;他还提出汉语照应语延续性强弱依次是零形式、代词和名词(同上:111-114)。延续性实际上是同指的照应词(代词、名词和零形式)之间的距离①。照应词的使用受人物、情节、时间、连词和结构等的制约(同上:130-137)。本章以下具体分析"他"在翻译和原创文学语料中的使用特征:其一,"他"的复现率。其二,多"他"同句的频率及特征,包括 1)"他"的指代能力;2)多"他"同句;3)不同指"他"交替使用等。

13.3.1　"他"的复现率

本章的"复现率"指两个"他"之间相隔的距离。这个概念是受回指理论的"延续性"影响产生的(徐赳赳 2003)。所谓延续性实际上就是同指的照应词(代词、名词和零形式)之间的距离。篇章语言学是通过"回数法"来计算同指词之间小句的数量。这种方法适用于小规模的篇章分析,但对于上千万字词的语料库而言,人工回数小句是不现实的。故本章用复现率的概念来计算代词和名词的"延续性",具体方法是:复现率=语料库的句子数÷"他"的实际频率。复现率的意义是衡量语料库中"他"与"他"之间平均相隔多少个句子,也就是各个"他"之间的平均距离。复现率的值越大,代词之间相隔的距离越远。为了比较翻译文学语料中人名和"他"的复现率,我们分别计算了各语料库中"他"和人名的复现率,见表 13.2②:

表 13.2　各语料库中"他"和人名的复现率

	原创汉语	翻译汉语	原创英语	翻译英语
句数	30070	39963	41159	36107
他(he/him)	6582	13092	12719	13577

---

① 具体方法是用采用"回数法"数出指同代词或名词(token)之间的小句数量(value),再用该值除以代词或名词的数量得到平均每隔 n 个小句出现的代词或名词数(徐赳赳 2003:111-114)。

② 本表中的句数和频次皆为语料库的实际统计数。其中"他"不含"他的"和"他自己";"he/him"为"he"&"him"频次之和,不含"his"。

（续表）

	原创汉语	翻译汉语	原创英语	翻译英语
复现率	4.57	3.05	3.24	2.66
人名	19671	12212	—	—
复现率	1.53	3.27	—	—

从表 13.2 我们可以看到，原创汉语文学语料中的“他”的复现率为 4.57，其含义为：平均每隔 4.57 个句子会出现一个“他”。翻译汉语文学中“他”的复现率为 3.05，即约每隔 3 个句子出现一次。显然，原创汉语中“他”的平均距离明显大于翻译汉语。另一方面，一般认为汉语倾向于使用“名词复现”的指代方式，为了证实这个看法，我们也统计了汉语语料库中人名的复现率。结果发现：翻译汉语文学中人名的平均距离(3.27)明显高于原创汉语文学(1.53)。简言之，语料统计表明：翻译汉语文学中“他”更为密集，人名则更为稀疏。

由于“汉英对应语料库”由英汉双向的原创和翻译语料构成，故我们也以同样的方法计算了英语语料中的“he/him”的复现率。可以看到，翻译英语中“he/him”也比原创英语更为密集。这是否能在一定程度上说明“显化”是翻译汉语和翻译英语的共同特征，还有待进一步考察。应当指出的是，上述计算方法只能非常粗略地比较“他”或人名在语料库中的平均距离，而未考虑原创汉语和翻译汉语的句子长度。对于具体篇章的“他”更适合用延续性来考察其回指特征。

13.3.2 多“他”同句

同一句子中出现多个他的现象在现代汉语中并不少见，但一般而言，汉语更倾向于用“零形回指”或名词复现的方式来进行照应。先来看多“他”同句的频率，我们对原创汉语文学和文学翻译语料中 2～6 个“他”在同一句子中共现的频率进行了统计：

表 13.3 多“他”同句的频率

	原创汉语		翻译汉语	
	原始频率	每百万词	原始频率	每百万词
总频率	7666	13658	15447	20445
2“他”同句	1708	3043	4202	5562
3“他”同句	377	672	1128	1493
4“他”同句	87	155	311	412

（续表）

	原创汉语		翻译汉语	
	原始频率	每百万词	原始频率	每百万词
5“他”同句	26	46	93	123
6“他”同句	10	18	29	38

由表13.3可见，汉语翻译文学语料中同一个句子中多个“他”连用的现象比汉语原创文学高很多。2～6“他”同句的情况，翻译汉语几乎都是原创汉语的两倍多。因此，相对来说，同一句子中多个“他”并用的现象在翻译文学语料中是较常见的现象如例(1)，而汉语通常采用人名与“他”交替使用的方式如例(2)。而例(3)这样同一句子中出现多个音译人名的情况，在翻译文学语料中出现的情况很少，该句读起来非常吃力，人名音译加重了读者的阅读负担。

(1) 那顶帽子使他很是为难，他刚想往上衣口袋里塞的时候就被前面的人夺了过去，那动作既轻巧又自然，不由得这位笨拙的年轻人心里不感激。“他能体谅人，”他想，“他准会帮俺帮到底的。”(《马丁·伊登》①)

(2) 但他没有走远，他只是在一个别人不注意的地方躲起来，他要跟着红艳去找王新——他断定红艳一定要找王新的。

(3) 波里娜·盖奥尔吉耶芙娜虽然没有向奥列格说明刘季柯夫叫他和杜尔根尼奇去的是什么地方，但是杜尔根尼奇立刻猜到，这是通往游击队的道路。(《青年近卫军》)

13.3.3 不同指多“他”同句

值得注意的是，如果一段篇章中有多个人物交替出现，汉语通常都会以人名来指代，一般不会在同一句子中同时使用多个不同指的“他”。然而，我们对翻译文学语料中多“他”同句的索引和分析中发现，翻译文学语料中存在不少不同指“他”在同一句话中出现的例句：

(4) 他$_1$背上的孩子会不会突然掉下来，他$_2$腰上的孩子会不会因为他$_3$父亲的动作打扰了他$_4$的美梦而大声啼哭？(《本性》)[他$_{123}$同指父亲，他$_4$指小孩，两个不同指“他”交替]

① 以下部分例句来自“当代汉语翻译小说语料库”(CCTFC)，参见胡显耀(2007)。

(5) 他$_1$ 作为军事审判官决定的事，只有作为文职特派员的他$_2$ 才能改变。(《九三年》)[他$_1$ 他$_2$ 各为两人，两个不同指“他”交替]

(6) 像这样被拖上街以前，别的孩子就不止一次大声喊他$_1$，讥笑他$_2$ 父亲，讥笑他$_3$ 老是当众宣扬他$_4$ 的宗教信仰。(《美国的悲剧》)[他$_1$ 他$_2$ 指儿子，他$_3$ 他$_4$ 指父亲]

(7) 他$_1$ 特别指给他$_2$ 看一个胖子，这人脸上流露出天真和善的神情，懒洋洋地倚在他$_3$ 身边一个女伴的肩上，用充满父爱的眼神深情地注视着他$_4$ 的女伴——一个淳朴的年轻姑娘。(《娜娜》)[他$_1$ 他$_2$ 和他$_{34}$ 分别三人，三个不同指“他”交替]

(8) 他$_1$ 又一次挥动左手，击中了那个炮手，炮手向他$_2$ 扑过来，抓住了他$_3$ 的上衣，扯下了他$_4$ 的袖子，他$_5$ 往他$_6$ 的耳朵后面狠狠揍了两拳，接着在他$_7$ 把他$_8$ 推开的时候，又用右手把他$_9$ 击倒在地。(《海明威短篇小说选》)[共出现 9 个“他”，他$_{1\text{-}4}$ 同指一人，他$_5$ 到他$_9$ 很难分清所指。]

我们在汉语原创语料库中未发现上述这种同一句子中多个不同指的“他”交替使用的情况。从汉语的角度来看，这种不同指“他”混用容易造成语义含混，让读者分不清“他”的所指。我们认为，这种现象可能有三个方面的原因：一是原文如此，英语等语言的句法结构层次清楚，因此出现在主句和从句中的 he/him/his 很容易从句法形式上得到区分，故不同指的 he/him/his 同句也不会造成歧义。二是由于翻译文学语料中人名音译复杂冗长的客观情况，译者在译文中无法采用人名复现的方式来明确代词所指。三是在使用“他”还是使用零形回指两种选择中，多数译者选择了“他”的复现，这体现了翻译显化这一基本取向，因为“他”的复现至少相对与零形回指来说更为明确一些。然而，译者没想到的是不同指的“他”交替使用反而造成了译文的含混和歧义。因此，我们认为，人称代词“他”的不同指同句是一种变异的翻译操作。

13.4 小结

综上所述，翻译文学语料中各种人称代词，尤其是第三人称代词“他”的使用频率显著高于原创文学语料。与英语原创小说相比，英语翻译文学语料除了 it 和人称代词的所有格频率较低外，其他人称代词的频率都呈现出与英语原创文学语料相似的高频率。人称代词在翻译文学语料中的复现率显著区别于原创文学语料和汉语总体，“他”之间的距离相隔更短。“他”在翻译文学语料中的照应功能显著增强，明显不同于原创文学语料更多使用“零形回指”的趋势。人称代词的增多和不

同指的"他"交叉使用做法客观上产生了变异的效果。

本章认为，人称代词显化的原因可能与英语等外国文学（主要是小说）的性质有关：一是人称代词所指代的人物的姓名译为汉语一般采取音译的方式。在汉语读者看来，外国人名显得冗长怪异。翻译中若让这些冗长和怪异的姓名不断"原词复现"显然是自找麻烦，降低了小说的可读性。第二，对英语等欧洲语言来说，人称代词，尤其是第三人称代词的语篇照应功能很强，人称照应是语篇衔接的重要手段。从欧洲语言翻译入汉语，第三人称"他"明显多于其他人称，可能也反映了这一特征。因为第三人称代词除了具有指别的功能外，还有第一、二人称代词所不具备的照应功能。

然而，为什么翻译文学语料中没有使用汉语常见的"零形回指"而是选择了反复使用第三人称代词呢？这种"他他"不已的现象是出于怎样的翻译操作规范？对译文又造成了什么影响？我们认为汉语翻译文学人称代词的差异性特征反映了现代汉语文学（小说）翻译的显化和变异的操作规范，即外译汉的小说翻译者在汉语的意合特征与欧洲语言的形合特征两者之间的实际选择是倾向于语法显化或"形合"。语法显化的总体目标是通过语法关系的明确化，降低文本阅读难度，增强翻译文本可接受性。但从汉语翻译文学的实际结果来看，显化的结果未必一定能降低翻译文本的阅读难度，有时反而出现了对汉语语法规范的偏离和变异。

第 14 章

基于语料库的翻译文体研究

现代文体学的发展为翻译研究带来了新的启示，本章从文体与翻译之间的关系入手，在前人研究的基础上，尝试建立基于语料库的翻译文体学研究框架，即借助语料库软件工具，采用语际对比和语内类比相接合的模式，从语言形式和文学叙事两个视角对各类翻译文本的文体特征进行量化描写、分析和解释，发现翻译文本不同于一般文本的原因，进一步认识翻译的本质。

14.1 引言

文体学与翻译研究均拥有悠久的历史，但现代文体学发展成为一门有影响力的独立学科大约是在 20 世纪初，翻译研究开始作为一门独立学科大约是在 20 世纪 70、80 年代。但文体学与翻译研究相结合却是近年来才引起了学者们的关注。申丹(1995:1)在讨论文学文体学与小说翻译的关系时指出，文学翻译这一领域本应十分适宜进行文体研究，但将文体学应用于文学翻译的尝试至少就英汉两种语言来说较为少见。Boase-Beier(2004)指出，多年以来文体学研究成果对翻译研究的影响出奇的小，翻译研究中的见解也较少带有文体学的因素，造成这两个领域缺乏互动的原因在于：第一，文体学研究的单语性质；第二，近年来的翻译研究更多地关注于文学和文化研究所带来的启示。随着语料库研究范式的建立，译作取向(translation-oriented)的翻译文体研究逐步发展了起来，有代表性的包括 Bosseaux (2004)、Millán-Varela (2004)、Baker (2000)、Olohan (2004)、Marco (2004)、Malmkjær(2003,2004)等。本章根据这一领域现有的成果，结合文体学与翻译研究的发展，尝试建立一个基于语料库的翻译文体学研究(corpus-based translational stylistics)框架，目的在于将语料库文体学(corpus stylistics)、基于语料库的翻译共性(translation universals)和译者文体(translator's style)研究结合起来对翻译文本进行考察，拓展语料库翻译研究的视角，探索翻译的本质。

14.2 文体的定义及文体学的研究范围

Leech & Short(1981/2001:10),将文体(style)定义为"特定语境下,特定人物出于特定目的对语言的使用方式",他们(同上:38-39)对这一概念做了进一步说明:1)文体是语言的使用方式:2)文体属于言语而不是语言的范畴;3)文体是由从语言总存(repertoire)中所做的各种选择组成;4)对文体的定义是根据语言使用域(domain)来进行的(如特定作者在特定文类或特定文本中会做出何种选择)。根据Crystal(1999:323),文体是个体或群体在一定情境中对语言的独特使用。Wales(2001:371)将文体定义为"可察觉的独特表达方式"。这里需要指出的是英语*style*一词并不仅指汉语中的"文体",有时它还被用来指"语体、风格"等。文体实际上是一个较为宽泛的概念,研究者因为研究角度的差异常常偏重其不同的部分,如体裁、语体等研究。程雨民(1989:1-2)将语体定义为"同一语言品种(标准语、方言、社会方言等)的使用者在不同的场合中所典型地使用的该语言品种的变体"。他(同上:46)进一步指出:"语体的实质是一些在使用场合上有区别的同一变体的选择。"陈国华(1997:48)指出:"文体和语体是两个范畴,前者主要指书面语言的体裁形式,包括散文体、诗歌体、书信体,等等;后者指一种语言的各种变体,包括书面语、口语、地域方言、社会方言、个人特色语(idiolect),等等。"总体上,文体研究因为视角不同一般包括文体类型(genre)、语言变体(linguistic variety)和文本类型(text type)三个方面的研究。

*Genre*一词源自拉丁语*genus*,原意为"种、类"。根据Crystal(1999:132),*genre*指艺术作品可识别的范畴,在文学领域中可包括诗歌、戏剧、小说等一般概念,以及科幻小说、犯罪小说和浪漫小说等次类概念;用于抽象义时可以指口语或书面语中任何可区分的变体,如歌曲、布道或对话。本章根据前一种释义将*genre*译成"文体类型",简称"文类"。这一类别是从文学研究的视角出发,按照语言表现形式从宏观上将文本划分为文学和非文学两个范畴,前者包括诗歌、小说、散文、戏剧等,后者则包括政论、传记、新闻等。语言变体是社会语言学术语,也称为语言变异或语言异体,Hudson(1996/2000:22)将其定义为"有相似社会分布的一套语项"。这里"相似社会分布"是指相同的社会环境或具有相同社会特征的语言使用者。语言变体可能与一种语言中的发音、语法或词汇选择等因素有关,也可能与语言使用者的社会地位以及教育程度等社会因素有关,还可能受社会语境或使用场合的正式程度制约。语言变体常被笼统称为语体,实际包括使用变体(use-related variety)和使用者变体(user-related variety)两个方面,根据语言的使用可将语言变体划分为语体和语域两个方面,前者是根据语言使用场合的正式程度,后者则与社会语

境，如职业、主题、谈话对象等因素有关。

Trudgill(1983:102)指出，原则上语体和语域是独立的，例如足球语域可与正式语体共现(如高级报纸的报道中)，或与非正式语体共现(如酒吧里的谈话中)，但许多时候两者往往是联系在一起的。根据使用者划分的语言变体称为方言，具体包括地域方言、时代方言、社会方言、标准语和个人特色语等。以上这些均可作为语言变体研究的对象。此类别是从语言学视角，从微观上按照语言的使用情境或正式(或礼貌)程度所进行的分类。根据 Hatim & Mason(1990/2001:140)，文本类型作为一个概念框架可以帮助我们按照服务于总体修辞目的的各种交际意图来对文本进行划分，这些意图至少包括四个方面：1)引起读者的注意；2)宣布某一话题；3)表达对某项计划的支持；4)通过论证说理。据此，Hatim & Mason(1990/2001:153-158)将文本划分为论说性、说明性和指导性三大类。但需要指出的是，这里文本类型的划分与文体类型划分在有些方面存在部分重合，如广告从文本类型上属于指导文的范畴，从文体类型上则属于非文学文本的范畴，电影字幕从文本类型上属于叙述文的范畴，从文体类型上则属于文学文本的范畴。

Trosborg(1997:16)指出，文体类型和文本类型的划分理论基础不同，但两者都是有效的文本契约，对前者的定义和划分是以系统的非语言学标准为基础，而后者则是以认知范畴或严格的语言学标准为基础。文本类型有时也被笼统称为文体，从写作修辞的视角，按照信息呈现方式可将文本划分为叙述、描写、说明、论说和指导五个类型。

Bally 认为，语言除客观表达思想外，常带有各种情感色彩，文体学的任务在于探讨表达不同情感特征的各种语言手段以及它们之间的相互关系，并由此入手分析语言的整个表达方式系统(申丹，2000:22)。申丹(2008:V)指出："文体学是运用现代语言学理论和方法来研究文体的学科，在某种意义上，它与语言学之间的关系是一种极为密切的寄生关系，新的语言学理论的乘胜和发展往往会催生新的文体学分支。"文体学的蓬勃发展也同样为翻译研究者提供新的研究视角，重新来审视与翻译活动相关的文体现象。对文体与翻译的讨论主要分为原作取向型(source-text oriented)和译作取向型两类。

14.3 翻译研究的文体学视角

14.3.1 原作取向型

原作取向型的翻译文体学研究主要包括两类：第一，从语文修辞角度探讨译文与原作在语文上的对应关系，代表人物如 Tytler(1907/2007)等。第二，从语言学

角度探讨译文与原作在形式上的对应关系，代表人物如 Nida & Taber(1969)等。

Tytler(1907/2007:9)早在18世纪末就明确提出，好的译文"写作文体和风格应当与原作的特点相同"，并且"译文应当具有原作的自如"。这里所说的文体和风格主要指原作的语文和修辞，也就是说，译文应当以源文本的语文和修辞为核心。Tytler(同上:63-64)指出，好的译者应当发现原作者文体的真正特点，确认其属于哪一类别，是严肃体、高尚体、随意体、活泼体、华丽体还是简单体，并将原作的这些特点在译文中加以明确体现。可见，Tytler 所讨论的"文体"实际上是"语体"问题，即语言的正式程度(degree of formality)。

Nida & Taber(1969:12)指出，翻译的本质是"在接受语中重新创造源语言信息最贴近、最自然的对应项，首先是意义，其次是文体"。根据 Nida & Taber(同上:13-14)，他们所说的文体不仅包括:1)文体类型，如诗歌(poetry)、散文(prose)等形式;2)文本类型或写作修辞模式，如叙述体(narrative)、说明体(expository)、论说体(argumentative)等;3)语体，如《圣经·马可福音》(Mark)的轻快活泼体、《圣经·路迦福音》(Luke)的优雅体等;4)修辞，如诙谐双关语(play on words)等。Nida & Taber(同上)认为，翻译中无论在意义还是文体层面，功能对等都是最重要的。

14.3.2 译作取向型

译作取向型的翻译文体学研究主要包括两类:第一，叙事学视角的文体翻译研究，代表研究如 Hermans(1996,2002)、Bosseaux(2004)、Millán-Varela(2004)等;第二，语言学视角的文体翻译研究，代表人物如 Baker(2000)、Olohan(2004)、Marco(2004)、Malmkjær(2003,2004)等。

Hermans(1996:27)提出，翻译叙事话语(translated narrative discourse)在文本中表现出的声音不止一种，其话语呈现方式也不止一种，他将叙事话语中包含的"第二位"声音称为译者的声音(translator's voice)，以此作为译者话语呈现的表征。Hermans(2002:11)指出，译作当中译者异质声音的痕迹随处可循，译作绝非是对原作的透明再现，而是具有多重、分散、混杂和复调(polyphony)性质的双重文本，旁注、脚注、括号说明以及序言等准文本形式无不表明两种不协调的声音共存于译作中，因此以"透明"来评价译作有些过于简单。这些研究主要是从翻译策略及翻译的社会表现形式探讨译者特有的存在方式，是从语言外部来观察翻译整体表现出的特征。Bosseaux(2004)考察了弗吉尼亚·伍尔夫《到灯塔去》(*To the Lighthouse*)及其三个法文译本中作为视点(point of view)概念语言学表现之一的自由间接引语(free indirect speech)的翻译。从叙事学视角考察了译者的话语呈现方式，探讨了译者的选择会对叙事结构的转换产生何种影响。Millán-Varela(2004)利用巴赫

金的对话理论(dialogism)和话语混杂(heteroglossia)等概念集中关注了詹姆斯·乔伊斯作品的加利西亚语(Galician)翻译,讨论了译者的声音是如何与源文本中其他声音相互作用,以及译者的声音有何种塑造功能。Millán-Varela(2004:52)在结论中指出,对翻译文本的语言学和文体学分析可以为研究者提供关于翻译主体及其意识与动机的关键性信息。

语料库翻译研究模式在此领域的应用将译者以语言和非语言形式所表现出的风格均作为研究对象,扩大了文体研究的范围。Baker(2000:245)指出,译者文体包括译者的选材、惯用的具体策略,包括前言、后记、脚注、文中词汇注释等,尤为重要的是个性的表达方式,即典型的语言运用习惯。可见,Baker 所谓的译者文体是一个较为宽泛的概念,而且这些特征是译者在翻译过程中的无意识产物,并且在一定时间范围内保持相对稳定。Baker(2000)利用类符-形符比、平均句长以及词项 *say* 的使用,对比了 Peter Bush 与 Peter Clark 两位译者的译文后得出结论:既然从理论上识别某个文学译者的文体是可能的,就有可能借助此类描写来描绘该译者选择再现的世界,帮助研究者将语言习惯与译者的社会文化处境(包括文化之间的关系、预期读者等)结合起来。Olohan(2004:153-167)通过"个体译者的缩略词模式调查"和"应用关键词分析探究译者的词汇选择"两个个案介绍了译者文体研究的具体操作,她同时指出,影响译者文体的因素包括文化中介的需要、读者的期望、意识形态以及编辑与修改等。Marco(2004:74)从 Leech & Short(1981)对"文体"的定义入手强调,文体并非仅是个体性的,也可以是集体性的,即特定文类或特定话语中标准化了的语言使用方式,也就是说,文体是由某一个体作者、某一特定作品等所做出的一整套(常常是模式化的)语言学选择。Marco 给出了一个建立在系统功能语法基础上的译文导向型文体分析模式(a model for translation-oriented stylistic analysis),从原作文体对译作的影响和译作文体的语言学表现两个互补的视角讨论了文体与翻译的关系,具体包括原作文体特征,以及个体或群体译者的文体特征如何在译作中分别得以体现两方面。Marco(同上:89)最后指出,文体与翻译界面尚有很多值得研究的问题,未来的实证研究应当解决如何对翻译文本中的问题模式加以分析,以及原作文体与以下因素的相互关系,这些因素包括:译者的个人偏好、两种语言的对比、目标语文学系统或特定群体或社会运动在诗学和意识形态方面的要求等,基于语料库的方法论在此领域大有所为。Malmkjær(2003,2004)以 H. W. Dulcken 对安徒生作品的英译为例对译者文体作了进一步探索,初步提出了"翻译文体学"(translational stylistics)这一概念,主要关注在原语文本既定的条件下,译者为什么会以特定的方式来塑造译文,对这些现象的解释除考虑语言因素外,亦可从语言外因素入手,如翻译规范、译语文本的目的等。

14.4 基于语料库的翻译文体学研究框架

近年来语料库方法开始广泛应用于文体学研究和翻译研究，其中语料库文体学(corpus stylistics)的一些研究成果开始引起学者们的关注(Semino & Short 2004;Wynne 2005;Hoover 2007等)，国内外学者所进行的基于语料库的翻译共性和译者文体研究均可视为翻译文体研究的一个方面，却未能就翻译研究中的各种文体现象建立起一套统一的基于语料库的翻译文体学研究框架。基于语料库的翻译文体学是以双语平行语料库为基础的描写研究，采用语际对比和语内类比相结合的综合对比的模式，通过计算机软件对各类文本(包括从文体类型、语言变体和文本类型角度分类的文本)的文体特征进行语言学的量化描写，在充分描写和分析的基础上从语言内、外加以解释。

在翻译共性和译者文体研究的基础上，翻译文体学研究的内容可包括两大类：1)语际对比文体(contrastive style)，即原文文体对译文的影响及其在译文中的表现;2)语内类比文体(comparable style)，是指翻译文本与目标语中具有类比关系(包括文体类型、文本类型、语域、时代等)的非翻译文本相比较，翻译文本整体上所表现出文体特征，具体又可分为三种：第一，目标语中特定时期翻译文本与非翻译文本的文体对比;第二，特定译者的翻译文本与其创作的非翻译文本之间的文体对比;第三，译出与译入翻译文本的文体对比。这些研究可以帮助翻译研究者认识到翻译文本不同于一般文本的具体特征及其产生原因，有利于我们进一步从文体视角认识翻译的本质。

翻译文体学研究的描写参数可分为语言形式参数和文学叙事参数(多限于文学作品，尤其是小说文本)两大类。

第一，语言形式参数，具体包括平均句长(average sentence length)、平均句长标准差(standard deviation of sentence length)、类符-形符比(type-token ratio)、标准化类符-形符比(standardized type-token)、词汇密度(lexical density)、关键词(keyword)等，这些数据均可通过WordSmith工具获得。不同类型文本的平均句长不同，根据平均句长和平均句长标准差就可以了解文本中各种长短句子的分布情况。类符-形符比指语料库中连续词语(running words)总数量与所使用不同词语数量之间的比值。类符-形符比取决于语料库的大小，根据Wordsmith Tools提供了一个标准化类符-形符比，可对不同长度文本或语料库的类符-形符比进行比较，做出关于其使用词语多样性程度的论断(Olohan 2004:80-81)。词汇密度是用来描写语料库中词汇词与语法词关系的比率，即词汇词占总词数的百分比。根据斯塔波(1996)，大多数书面语文本的词汇密度超过40%，而口语文本的词汇密度

变化幅度较大，书面语文本之间也存在差异，这通常与文体类型相关，文学文本的词汇密度要比学术文章等非文学文本低，表现出的趋势是信息负载逐渐增高（同上：81）。词汇密度常被作为计量简化（simplification）的一种手段（如 Laviosa 1998a，1998b）。Olohan（2004：81）指出，不同类型或不同语域文本组成的两个不同语料库的频次表，在文本类型或语域方面的差异会在一些词汇词或语法项的频次差异上有所体现，利用软件来比较两个语料库的频次表，就可以识别出较两个语料库的关键词。通过以上这些参数就可以发现源文本与目标语文本、目标语翻译文本与非翻译文本之间在词汇多样性、信息负载、主题焦点化和创造性词汇使用方面的差异，并对这些差异进行量化，认识翻译文本的文体特征。

第二，文学叙事参数，如从两个被比较文本中对人称代词和自由间接引语的使用来考察译者的选择会对叙事结构转换产生何种影响，以及翻译文本的叙事结构与非翻译文本有何不同。

可见，借助语料库工具的翻译文体学研究主要有三个特征：第一，从翻译文本入手，将翻译文本作为独立整体加以分析；第二，主要从翻译语言中来发现译者和翻译过程留下的痕迹；第三，不再将源文本-目标文本之间"对等"与否作为衡量译文的文体标准，而是作为参照文本来探讨源文本因素对目标文本的影响。

14.5 小结

翻译文体学研究是通过语言对比这一纽带将文体学与翻译研究结合起来，改变了翻译研究仅仅关注语言转换或文化研究的单一模式局面，而是在双语语料库基础上将语言对比、文体学研究与翻译研究结合起来，拓宽翻译研究的视野。将文体学研究对象的范围扩大到翻译文本，即以翻译文本为基础，从文体学的视角探究翻译文本所表现出各类特征的根源。目前，翻译文体学研究尚处于探索阶段，进一步的研究还需要建立系统的翻译文体学检测项，制定一套针对翻译文体行之有效的语料库标注符集，从而对翻译文体进行更为客观的描写和解释。

第3部分

语料库与翻译教学

第15章

语料库翻译教学的理论与原则

本章分析了对应语料库、翻译教学和自主学习之间的内在联系，在此基础上探讨汉英对应语料库在翻译教学中应用的原则和方式。研究认为，对应语料库可以为自主学习和发现式学习创建有利的实施环境，而自主学习和发现式学习有助于翻译教学效果的提高，这是对应语料库应用于翻译教学的理论依据。基于此，对应语料库在翻译教学中运用的基本原则是观察先于归纳、呈现先于讲解、学生自导先于教师指导。

15.1 引言

现代语料库是基于当代信息技术的真实语言材料处理、存储和检索系统。近20年来，基于语料库的语言研究取得了长足发展，与之相关的语言研究方法也得到了学界的广泛认同。对应语料库起步较晚，大都在20世纪90年代初期开始创建，但已展现出良好的发展和应用前景。在欧美，这类语料库有十多个，涉及十多个欧洲语种，较著名的有Hansard(法-英对应语料库)、ENPC(英语-挪威语对应语料库)(Véronis 2000:14-15)；在中国，已初具规模且拥有自主研发检索系统的对应语料库是北京外国语大学"汉英对应语料库"(王克非2004a;该库现有可检索语料2000万字/词)，也是本课题研究的一个主要平台。语料库的创建有两个主要用途：一是用于语言研究，二是用于外语教学。与前者相关的研究内容丰富，涉及语言对比、双语辞典编纂和翻译研究(Baker 1995,1999;Gaussier et al. 2000;Serpellet 2000;Hunston 2002等)。而第二个方面的研究尚不充分，且大都基于单语或学习者语料库(如Sinclair 1996;Wichmann 1997;Hunston 2002等)。

将对应语料库应用于翻译教学的研究刚刚起步(Bowker 2003;Bernardini 2004b)。近年来，国内已有研究关注语料库在翻译教学中的应用(柯飞2002a,2002b;郭红2004;于连江2004;王克非2004a,2004b等)，但这些研究还没有解决平行语料库与微观翻译教学环节的衔接问题。主要原因有：从创建目的上看，对应语料库主要服务于语言对比、辞典编纂和翻译研究，而非外语教学；从创建技术上

讲，目前常见的对应语料库多为词性标注，这就影响着语言信息的提取深度，还无法同时提供翻译所需的语义-句法信息；从检索技术上说，语料库使用者很难一次检索到适合课堂使用的语言材料。总之，关于对应语料库在翻译教学中的应用在语料处理、检索方法以及与翻译课授课内容的结合上还有很多问题有待解决。本章的主要内容是：汉英对应语料库在翻译教学中应用的理论依据；汉英对应语料库在翻译课堂教学中应用的原则；汉英对应语料库在翻译教学中运用的主要方式。

15.2 对应语料库在翻译课堂教学中应用的理论依据

根据目前的研究，对应语料库在翻译中的应用价值主要体现在以下几个方面：1)语料库可用来提高一般意义上的语言意识(Hunston 2002:123)，即译者对两种语言的驾驭能力，所以在翻译培训和集中探讨某些翻译技巧问题时非常有用；2)通过对比语篇观察两种语言是以何种方式来编码对应意义的(Barlow 2000:111)；3)可使用双语语料库数据分析为译者提供不同语言中可能的翻译单位或者对等篇章单位的集合，提高译者的翻译意识，即对翻译难点的处理意识(awareness)和对翻译策略的运用意识，有助于改进翻译的终端产品(Serpellet 2002:359-360)；4)双语语料库还有助于翻译对等的研究(Leech 1997:22)，帮助研究者洞悉翻译的本质(Hunston 2002:128)。

以上论述虽未谈及翻译课堂教学，但对应语料库对于翻译学习者和翻译从业者的辅助作用是相似的，只是作用的方式和程度有别。Barlow 较早探讨对应语料库在课堂教学中的应用。他指出，在课堂上使用对应语料库能使语言学习者直接探究(可能是应老师的要求)两种语言中词和结构的主要对应项，帮助学生在面对这些信息的情况下建立新的图式并构建恰当的图式-意义关联(Barlow 2000:113)。而 Bernardini(2004a,2004b)的研究表明，对应语料库的使用有助于发现式学习环境的创建。

我们认为，将对应语料库应用于翻译教学有利于创造自主学习(autonomous learning)的环境，是自主学习理论的自然延伸。正像对应语料库在翻译研究和翻译实践中具有重要价值，对应语料库在翻译教学中也起着重要作用，有助于翻译教学和技能培养水平的提高。

15.2.1 语料库与自主学习

学习自主性就是学生负责自己学习的能力，这一理念通常被认为是推行自导式学习(self-directed learning)的结果。自主学习这一理念出现在 20 世纪 60 年代，时间不短，但它具体指什么，现在还没有完全一致的看法。

一种看法是完全自主，即自主学习和课堂学习无关。目标、进度和学习效果的评估完全由学习者自己确定(Benson 2001:8)，由学习者自己负责与学习相关的决策并实施这些决策(Dickinson 1987:11)。

另一种看法是自主学习可以在课堂教学环境下实施，即改变传统教学的基本架构，将教学过程的控制权从教师移交给学生。目前的研究显示，培养学习者的反思和分析能力是提高学习自主性的核心，而这一能力的发展有赖于学习者全心投入到人际互动中去(Benson 2001:14)。依照这种学习观，那种将自主视为纯个体学习行为的观念需要纠正，自主学习并非只是课堂教学之外的学习活动。即使在传统教学模式下的课堂里，只要有合适的合作和互动环境，也同样有利于学生发展自主学习能力，实现预期的教学目标。

如果我们同意后一种自主性学习观，对应语料库就可以成为翻译课堂教学中自主学习实施的理想平台。从技术上讲，语料库可为学习者快速提取学习材料提供方便；从内容看，适当规模的语料库可为学习者提供丰富多样的例证。

较之传统的翻译教学模式，语料库支持下的翻译教学更易于加强学习者学习的主动性和自主性，因为这时学习者可以是发现者，甚至是研究者。如果把学习者视为发现者，就应该引导学习者去观察和探寻外国语言现象，就像语言学家以同样的方式去发现此前未曾注意的本族语语言事实(Bernardini 2004a:23;2004b:102)；如果把学习者视为研究者，就应鼓励学生按照自己的选择去获得、组织并研究真实语言素材。这就为学生作为“研究者”去开辟新的研究领域带来了实实在在的希望(Leech 1997:10)。

本章更倾向于将学习者视为发现者。鉴于翻译教学的经验性和实践性特征，对应语料库为翻译教学提供了一种自主学习环境，学生可以从中发现母语和外语各自的特征以及转换的方式，培养起翻译意识，逐渐形成运用翻译策略的能力。具体说来，通过考察一个词或者一个结构的多个翻译例证，个别性的因素就会被规律性替代；而这些由学习者自己归纳出来的、有动因的翻译策略累积起来，就可使学习者感知到有效的、具有普遍性的翻译处理方式，并形成有意义的概括。

15.2.2 翻译教学中的自主学习

目前有研究显示，自主学习和翻译教学在理论上有很多契合点。

在课程设置上，以学习者为中心的课程设置与其他课程设置的关键差异在于，它强调教师与学习者之间的共同努力，因为学习者与涉及课程设置的内容和教法的决策过程是紧密相关的(李德凤、胡枚 2006)。对翻译学习者而言，以学习者为中心是学习者通过自主学习逐渐成为翻译工作者的过程。李、胡进一步指出，学习者学习翻译的过程并不单是从教师那里习得知识，而是在实践中累积建构自己的

专业知识，教师则只是发挥指导与协调的作用。

在实施原则上，王宇(2003)指出，所谓“以学习者为中心”的翻译教学，指的是从教学目标的设定、教学材料的选择，一直到课堂教学、作业评估等环节充分考虑学生的因素。具体表现在：1)让学生参与教学内容和翻译材料的取舍；2)组织有效的课堂讨论和小组活动以激发学生的参与意识；3)教学方法与手段应侧重于培养学生严谨的工作作风和批判性思维能力，不迷信任何参考译文或译者；4)评判学生译作时，在充分肯定其闪光点后，注重分析和挖掘误译或错译的原因等。

以上研究所使用的“以学习者为中心的翻译课程设置”是一种教学模式，与这一模式相关的教学设计实际上是围绕自主学习策略进行的(Yang 2006)。这要求在翻译课程设置上应考虑学习者的自主性、创造性和互动性，协调学习者、翻译教学和市场需求之间的关系。发挥学习者的自主性和创造性实际上就是相信并充分运用学习者观察、发现、分析和制定策略解决翻译问题的能力。互动性主要指学习者之间或师生之间及时反馈各自的观点和发现，并通过协作和商讨来解决问题。

15.2.3 真实语料与翻译教学

对应语料库存储的是一个双语或多语文本，这类语料库为翻译教学提供了一个理想的实施平台。对应语料库应用于翻译教学的优点主要体现在以下四个方面：

第一，与传统基于直觉的分析相比，基于语料库的教学有助于学生发现语言使用的复杂性和语言中的型式(pattern)(Barlow 2000：106)。

第二，学生花些时间审读真实篇章中有语境的例证，这样要比完全或者主要依赖语法书和教科书中那些孤零零的范例更可取(同上：106)。

第三，可以使用语料库将某些词或语法特征从一种语言里挑出来和另一种语言里的对应项进行对比，查找异同。这样，对应语料库就可以用作一个专门的参照系统，把学习者的注意力转向成熟译者或专家译者所发现的典型问题的典型(或非典型)处理方法(McEnery & Oakes 1996：211)。

第四，如果把翻译教学看作文化顺应过程，即学习者愿意把自己作为一个某一语言社团的成员，并按照这一社团所共享的思维、行为和交际模式来有效地传递信息，对应语料库的实用性就更为突出(Bernardini 2004a：20)。

总之，对应语料库(包括翻译语料库)已经成为翻译教学的重要工具(同上)。按照语类或题材分类存储的语料库可以使学习者对比原语和译语语言、原语和译语语篇，甚至是两种语言的同类语篇。这就给发现式学习提供了丰富多样的学习材料和材料获取渠道，满足翻译学习者自主学习的需要。

15.3　汉英对应语料库与教学内容的设计

传统翻译教学内容的编排是以结构为纲，大都从词素、词的翻译谈起，然后扩展到词组、句子和篇章层面的转译上（陈定安 1995；陈宏薇 2004；张培基 1980 等）。这种编排的优点是层次分明，编排条理，也易于授课时使用；缺点是教师不好把握授课内容的难易程度（只能假定语言单位小的处理起来更容易些，但这仅仅是一种假定，还没有人能证明词或者词组一定比句子更容易处理），不易确定教授内容是否具有普遍性，更不便提供丰富多样的译文。总之，这样的翻译教学规定性更强，不便有效培养学习者处理实际翻译问题的能力。

基于对应语料库的教学内容可以按照频率原则编排。对应语料库可以帮助教师检索出有关某一语言现象的常用翻译技巧和策略，频次高的先讲，低的后讲。这样可以引导学生先掌握有普遍意义的技巧，在此基础上探究和掌握频次低、但难度更大要求更高的翻译手段和处理策略。另外，虽然英汉翻译和汉英翻译使用的翻译手段大致相似，但各种翻译手段的使用频率也会因目的语不同而不同。这时，就更需要语料库支持才能确定哪些技巧和策略应为优先安排的教学内容。

使用语料库可以集中而且方便地展示大量的翻译实例和素材，这一点用纸介的教材也可以做到，但很不经济，无法灵活掌握呈现语料的量。

15.4　汉英对应语料库运用于课堂上的可能方式

有关语料库在语言学习过程中的应用研究大多是基于单语语料库。近年来，已有研究注意到语料库在翻译教学中的应用（封一函 2001；柯飞 2002a，2002b；王克非 2004a，2004b；郭红 2004；于连江 2004 等），研究涉及的理论问题和可能的应用方式都有很高的参考价值，但大规模对应语料库在翻译课堂教学中究竟如何使用，目前还没有明确、具体的实施办法。所以，有必要为此提出可行的实施思路。本章认为，基于自主学习原则、基于语料库所提供的检索技术支持、基于翻译教学的实践性特征，对应语料库在翻译课堂教学中的应用可以按照下列步骤来实施。

第一，语料呈现——刺激：使用平行语料将语言学习者的注意力集中于关键词/结构的翻译对应项。语料由教师挑选，应学生要求检索，或学生自行检索。

第二，学生浏览语料：通过考察语料，学生自行建立新的形式-意义关联。

第三，学生报告观察心得并记录自认为有意义的发现：通过观察语料可以经验性地展示（或让学生自我发现）两种语言对应的一对多、多对多性质，使初学者建立起语言意识识和翻译技巧意识。教师及时得到反馈并回应反馈。

第四，事前和事后测试：检验语料呈现的效果和学生对语料的敏感程度。

第五，教师和学生共同讨论与所呈现语料相关的翻译策略。

在这种教学环境下，学习者的观察结果可能是局部或者经验式的概括，但这一观察结果可以通过新的证据来检验、修改和丰富。从外语学习和翻译技巧意识上看，学习者通过观察双语对应语料可对搭配、语义偏向和语义韵律等现象更加敏感；翻译技巧和策略可以通过学生间和师生间互动讨论逐步形成，这有助于学生形成稳定、持久和灵活的翻译策略。而在传统课堂环境下，单靠教师课堂传授翻译技巧很难做到这一点。

15.5 语料库翻译课堂教学对资源的要求

15.5.1 语料和软件资源

英语单语语料库（如 BNC）可作为参照库，用于检索、检验并确定汉英翻译中可能使用的结构和词汇用法是否合乎目的语规范；其他网络资源如各种搜索引擎都可以充当有效的参考资源。

双语语料库（如"汉英对应语料库"）可用于检索对译索引，提取课堂教学需要的翻译素材；该语料库主要由教师控制使用，教师可以根据学生的要求及时提供内容丰富的对译索引。学生可以通过观察丰富和多样的对译语料来讨论、归纳并总结出翻译技巧。

翻译记忆系统（如 Trados）的特点是不直接提供翻译，而是根据匹配率提供词汇、词组和句子的汉英参考对译（当然还可以是日语、德语、法语、俄语、西班牙等语言之间的互译），翻译学习者根据系统所提供的对译语料调整或修改自己的译文，这也是给学生提供自我监测和反省自己翻译行为的机会。更重要的是，该系统的优点是可在安装在计算机上供学习者根据自己的需要随时调用，也可以安装在小服务器上供学生课外自修时使用。

15.5.2 多媒体教学环境

课堂教学使用计算机网络教室进行语料库辅助的翻译教学有以下优点：

第一，可以直接进入教师专用的服务器获取预先设定的部分教学资源。主要由教师根据课程进度和教学计划提供有针对性的翻译例证，学生可根据需要自行决定提取和观察的数量。

第二，学生可以使用电脑上的翻译记忆软件来为翻译提供参考资源。

第三，学生可以通过共享程序和邮件，或小组讨论来交流翻译信息，通过相互

协作完成翻译。

第四,教师通过控制台提供专家译文供学生观摩、评价、学习并选择恰当的翻译方法。

15.6 小结

传统翻译教学理论注重技巧传授和学生的实践能力,教师在教学过程中起主导作用。在这种教学理念下,学生面对的往往是教材中数量有限且少有例外的零散范例,翻译质量的评估几乎完全依赖教师的直觉判断,学生的自主判断和自主构建策略的能力因之难以充分发挥。本项研究对这一传统理念做出修正,认为翻译教学的目标是提高译者灵活运用翻译技巧来处理问题的能力,因此,分析真实文本显然有助于学习者将来承担好真实的翻译任务;自主学习环境是提高翻译教学效果的关键,而语料库是实施自主学习的重要物质条件。这样,在语料库支持下的发现式(自主式)教学环境里,翻译教学就可以从技巧传授转向技巧意识培养,还可以丰富评估的方式。这不仅意味着传统翻译教学理念的改变,也意味着学生学习观念的改变。

本章从理论上论述了对应语料库用于翻译教学的诸多优点:1)避免翻译学习者盲目接受技巧和译文分析的框架,在掌握典型翻译技巧的同时注意到翻译的多样性;2)最大限度地挖掘合作性学习的潜能;3)为学习者的情感投入和自我评价提供机会,提高学习者自我投入的积极性;4)有助于学习者积极地看待翻译,独立思考并愿意寻求解决问题的方法。

第 16 章

双语对应语料库与翻译教学

本章探讨平行对应语料库在翻译教学中的应用价值。我们认为，双语对应语料库对于翻译教学是有用的参考工具或工作平台，其作用主要体现在：1)对某一检索词或短语提供丰富多样的双语对译样例；2)为常用结构提供多种双语对译样例，供讲解和仿习；3)提供丰富的可随机提取的一本多译作为对照参考。在翻译教学中利用双语平行语料库和检索工具，其作用是其他教科书和工具书所不能替代的，在翻译教学中有广阔的使用前景和潜在的开发价值。

16.1 引言

翻译教学离不开双语语料，而大规模的双语对应语料库也是在近十几年里开始研制，经过理论探讨和实证检验，已经逐渐形成了基于语料库的翻译研究范式，但在翻译教学方面对双语库的应用尚属一个新的探索领域。凭借这种双语语料库，不仅可以进行语言对比研究，还可利用它开展广泛的翻译教学与研究(包括自动翻译)，以及双语词典的研编。长久以来，翻译教学的重点一直放在探讨译文与原文之间如何达到对等的问题上，这种以"对等"为核心概念的教学模式通常都是以独立的原文文本与参考译文对照、讲练为手段，缺乏对学生翻译意识的培养，抹杀了学生在翻译训练中的主动性和创造性。本章以汉英对应语料库为语料来源，以大量例证来探讨双语对应语料库在翻译教学上的用途。

16.2 双语对应语料库辅助翻译教学及其特点

所谓双语对应语料库辅助翻译教学，是指借助大规模原文及其对应译文的电子文本，辅以计算机统计手段的翻译教学模式，该模式旨在从翻译产品入手，通过观摩、对比、分析、借鉴的方式发挥学生的主动性，培养学生的翻译意识，在评估他人译作和自我实践的基础上提高学生的翻译技能。目前国际上关于利用双语对应语料库提高翻译教学的研究已有一些，如 Zanettin, F., S. Bernardini & D.

Stewart(2003)所编《翻译教育中的语料库》(*Corpora in Translator Education*)论文集,其中不少论文有一定参考价值。Bernardini(1997)指出,翻译教学辅之以对应语料库检索,便于翻译专业学生形成一种翻译的"意识"、"反射"和"应变",这些技能"使专业译员有别于那些不熟练的业余爱好者"。Zanettin(1998)也举例证明双语对应语料库应用于翻译教学有助于提高学生理解原文并用更流畅的目标语进行翻译的能力。Friedbichler & Friedbichler(1997)的研究表明,将外语译成母语时,参照双语对应语料库核对有关翻译问题,能够帮助翻译学员或译者提高表达能力,减少译文的错误。将母语译成外语时,对语料库工具的需要更是成倍增加。利用对应语料库(配以合适的检索工具)便于查找特定表达方式的译法,使术语和短语的翻译更加准确、地道,而且常常可以提供多个翻译选择或翻译参考,比双语词典的例证更为丰富、更为真实。另一方面,双语对应语料库也是开发机器翻译系统、改进翻译软件的主要资源,如基于例句的机器翻译(EBMT)和基于统计的机器翻译,都是以对应语料库为基础的。这些对今后的翻译教学都会有很大的帮助。国内因双语对应语料库的研制起步较晚,应用于翻译教学研究的论述较少,近年在这方面的讨论有主要有柯飞(2002a,2002b,2004)、叶常青(2003)、于连江(2004)、韦瑶瑜、雷蕾(2006)等。

H. L. Dreyfus & S. E. Dreyfus 将专业技能习得的过程划分为五个阶段:1)新手阶段(novice),识别与该技能相关的各种客观事实和特征,并在此基础上习得决策行为的各项规则;2)高级初学者阶段(advanced beginner),识别各种相关的情境特征,包括明确界定的和未加明确界定的;3)能力培养阶段(competence),依照目的,按照层次进行决策;4)效率阶段(proficiency),直觉的理解加上有意识的行为;5)专业技能阶段(expertise),理性、自发行为的产生(转引自 Chesterman 1997: 147-150)。从这一过程可以看出,任何技能的习得过程都会经历由个体认识到整体认识、由有意识的反映到无意识的反应、由分析性决策到直觉性决策的过程(同上:150),翻译能力的习得也不例外。就学生翻译意识的培养,Chesterman 有三个建议:第一,提供给学生各种类型的翻译文本,将原语文本与翻译文本加以比较,来认识各种原语-译语关系;第二,将某一原语文本以不同的方式进行翻译,来认识原语-译语关系在具体翻译实践中的变化;第三,对翻译个案中不恰当的原语-译语关系进行研究,尝试提出新的策略(同上:157-158)。这些方法无疑在翻译教学实践中具有重要的指导意义,而双语对应语料库正是实现这些设想的较佳工具。

Monzó(2003)指出,要让翻译学习者认识翻译,最好的办法就是让他们置身其中(exposure),而双语库是创造此种翻译环境的一条途径,因为通过双语平行语料,不仅能让学生直接观察原语语言系统的规则,也能看到翻译文本的自身特征,

建立对翻译转换的感性认识。与传统的翻译教学模式相比，双语对应语料库辅助翻译教学具有三个特点：第一，在理念上，借助真实的翻译产品帮助学生建立关于翻译的意识。如对翻译教材的认识，不仅局限于以独立的词汇、句子、文本及其参考译文为对象，而是以大规模的各种原文及其译文的电子对应语料作为工具，以计算机作为辅助手段，来帮助学生认识与翻译实践相关的各种因素；第二，在性质上，改变传统翻译教学中规定性(prescriptive)方法，将描写性(descriptive)因素融入翻译教学实践中，通过观摩现有的翻译产品，进行对比分析，加深对于翻译活动的认识；第三，语料库辅助翻译教学的核心原则是让学生成为教学活动的主体，调动学生的主动性，在观察、对比、分析、借鉴中提高学生翻译的意识和应变能力，从感性认识达到理性和自发的反应。本文主要讨论双语对应语料库对于翻译教学有何用处，或如何将其用之于翻译教学。

16.3 双语对译检索用于翻译教学

北外的汉英对应语料库是一个句子层面上双语、双向对应的语料库，检索任何词语，无论英语或汉语，都可将该词语所在的所有语句提取出来，并同时显现该语句所对应译文，即双语始终链接，在一个界面上显现。在北京大学计算语言学研究所的帮助下研制的检索软件(concordancer1208)可以进行词语、特殊结构、搭配关系、词频等检索，包括英语单词的各种屈折变化形式可一并检索(如输入动词 be，可同时将 is，am，was，are，were，being，been 一并检出)。这样，用之于翻译教学，该语料库至少可以完成一下辅助工作：

一、对某一检索词或短语提供丰富多样的双语对译样例；

二、对常用结构的提供多种双语对译样例，供讲授者讲解，学习者借鉴；

三、提供可随机提取的一本多译作为对照参考。

下面我们分别以例证加以说明。

16.3.1 词语的对译检索

英语和汉语绝大部分词语都不是一一对应的，往往有多个义项和相应的译语，这在翻译中随时都能感受到。学生在翻译学习或翻译练习中宜加强这方面的意识。双语对应语料库能在检索时提供较多的例句，有助于学生丰富对译词语，感受对译语境。

16.3.1.1 overcome 与“克服”

例如随机抽取“克服”这个词进行检索。在汉译英非文学语料(85 万字词)中，我们检索到 24 处使用“克服”，英译时 15 处使用 overcome(含其屈折形式，下

同)，如：

(1) 有了这样的团结，任何困难和挫折都能**克服**。

With unity like that we can ***overcome*** any difficulty or setback.

(2) 要**克服**一个怕字，要有勇气。

We should ***overcome*** our fears.

但有九处作另外的处理。如：

(3) 那次大会总结了我国民主革命二十多年曲折发展的历史经验，制定了正确的纲领和策略，**克服**了党内的错误思想……

The delegates summed up the historically significant experience gained in the course of the twists and turns of that revolution during the preceding quarter of a century, formulated a correct programme and correct tactics and ***straightened out*** the wrong ideas inside the Party.

(4) 但是，我们的社会主义政策和国家机器有力量去**克服**这些东西。

But with our socialist policies and state apparatus, we shall be able to ***cope with*** them.

(5) 但是我们有信心，我们的党、我们的国家有能力逐步**克服**并最终消除这些消极现象。

But we are confident that our Party and our country are capable of gradually ***reducing*** these negative phenomena and eventually eliminating them.

(6) 抓住典型，处理了，效果也大，表明我们下决心**克服**一切阻力抓法制建设和精神文明建设。

Dealing with these cases will have the most effect; it will show our determination to ***surmount*** all obstacles in strengthening the legal system and promoting ethical progress.

(7) 第二个目标是**克服**官僚主义，提高工作效率

The second objective of political structural reform is to ***eliminate*** bureaucratism and increase efficiency.

(8) 要相互容忍，相互谦让，相互帮助，相互补充，包括相互**克服**错误和缺点。

You should be tolerant and generous with each other, help each other, complement each other's thinking and help ***correct*** each other's mistakes and

shortcomings.

我们从英译汉文学语料(185 万字词)中搜索到 18 个 overcome,只有三个对译为汉语的"克服"。从英译汉语料中反查"克服"所对的英文原文,查到 13 个"克服",也只有三处对应 overcome,其他的对应如:

(9) Knowing your natural temper better than I, he could the better judge what arguments to use, whether of tenderness or terror, such as might ***prevail over*** your hardness and obstinacy, insomuch that you should no longer hide the name of him who tempted you to this grievous fall.

由于他比我更了解你的秉性,他应该是个更合格的法官,他更清楚应该选用什么样的刚柔相济的辞令,来**克服**你的桀骜不驯以使你不再隐瞒那个诱惑你如此堕落的人的姓名。

(10) She ***conquered*** her fears, and spoke:

她**克服**了内心的畏惧,终于启齿了:

(11) But, few cared enough for that to think of it a second time, and, in this matter, as in all others, the common wretches were left to ***get out of*** their difficulties as they could.

可是注意到这类事件并加以考虑的人却很少因而在这件事上也跟在别的事上一样,普通的穷苦百姓便只有自行努力去**克服**困难了。

(12) Greater things than the Doctor had at that time to ***contend with***, would have yielded before his persevering purpose.

即使当时医生要**克服**的困难比现在还要大得多,在他那坚持不懈的努力之下困难也是会退让的。

(13) He, on this part, had pretty well ***fought*** his weakness down.

至于希尔弗,他已经比较成功地**克服**了自己一时的怯懦。

16.3.1.2 find 与"找/发现"

又如英语词 find。从上述汉译英语料中查得 86 个 find,由"找"转译来的仅十个,大部分为其他灵活的处理:

(14) 这条**路**还没有**开**出来,各行各业,包括企业,都要解决这个问题。

We have to ***find* ways** to provide such opportunities. People in all fields of endeavour, including those in enterprises, should try to solve this problem.

(15) 我们现在一方面是知识分子太少,另一方面有些地方中青年知识分子**很难**起作用。

While there is an overall shortage of intellectuals, in some places young and middle-aged intellectuals are ***finding* it difficult** to play a useful role.

(16) 我没有想到,我们党内有些同志也抽象地宣传起人道主义、人的价值等等来了。

I am amazed to ***find* that** some of our Party comrades are preaching humanism, the value of the human being and so forth in abstract terms.

(17) 有些话不一定准确,可能考虑不周到,但是要把世界局势稳定下来,总要**想些主意**。

Some of my remarks may not be precise or thoughtful enough. But we must rack our brains to ***find* ways** to stabilize the world situation.

(18) 即使资本主义发达国家之间相互比较起来也各有优缺点。

Even when we compare the developed capitalist countries, we ***find* that** each has both strengths and weaknesses.

(19) 南方不发展,北方还有什么市场。

For where can the North ***find* a market** for its products if the South remains underdeveloped?

(20) 余秋里同志告诉我,现在军队的养猪专业人员,一到地方就有用处,司机到地方是最受欢迎的。

Comrade Yu Qiuli has told me that soldiers trained in raising pigs can readily ***find* jobs**. And drivers are in great demand.

(21) 我们感到,虽然战争的危险还存在,但是制约战争的力量有了**可喜的发展**。

We now think that although there is still the danger of war, the forces that can deter it are growing, and we ***find* that encouraging**.

(22) 只要不放松,认真抓,就会**有办法**。

If we work hard, we shall ***find* a solution.**

(23)"消肿",比较难的是**安置**退下来的几十万干部。

But if we do cut back, it will be hard to ***find* jobs for** several hundred thousand demobilized army cadres.

(24) 解决这个问题**从何入手**,要很好研究。

We have to make a careful analysis to ***find out*** how to go about political reform.

(25) 这些人恰恰就在共产党里

They are to be ***found***, of all places, inside the Communist Party.

从上述英译汉语料里见到 448 个 find,许多并不对译为"找"或"发现";实际上,译成"发现"有 104 个,译成"找"的有 77 个,即大部分(267 个)为灵活对译:

(26) 只要汤姆讲的话有百分之二十是真的,她也就心满意足了。

and she would have been content to ***find*** twenty percent of Tom's statement true.

(27) 玛丽拿着他的书,要听他背诵,他就竭力地云来雾去地往下背。

Mary took his book to hear him recite, and he tried to ***find his way*** through the fog:

(28) 他半心半意拿起笔,不晓得是不是还想得起什么,能给他写在日记里。

He picked up his pen half-heartedly, wondering whether he could ***find*** **something more to write** in the diary.

(29) 不过,说老实话,我简直不敢相信他就是那同一个人呢。

But truly, forsooth, I ***find*** it hard to believe him the same man.

(30)"你咋知道?"

"How did you ***find*** that ***out***?"

(31)"你是说,若是认定他有罪的话?"

"If he's ***found*** **guilty**, you mean to say?"

从上述关于"克服/overcome"和"find/找/发现"等词语的对译检索,我们可以看到,它不仅能帮助学生查找合适的对译,还有助于学生提高英语的语感,特别是做汉译英时。例如需用英语表达"克服"之意时,学生就会不单是用 overcome,而能从语料库丰富的样例里学会选择更适其语境的词(如 surmount, cope with, conquer, prevail over, content with, fight down, get rid of, get out of, reduce, eliminate, correct 等);另一方面,从英语 find 的使用上也可体会出它远不只是我们所熟悉的"找"、"发现"之类意思(如(14)至(25)诸例),而这样的训练不是仅利用词典所能得到的,特别是通过英译汉语料反查汉语词语的对应英语词语。

16.3.1.3 条件连词 if

以单个文本作为翻译教学的材料,探讨翻译中对特定词类的处理,如汉语中表示条件关系的连词(如表 16.1 所示)在汉译英中是否都会以表示条件关的英语连

接词(如表16.2)所示。

表16.1 汉语中表示条件关系的连词

(陆俭明、马真1985;黄伯荣、廖序东主编2002等)

连接关系类型			前置连词	后置连词	连接副词
条件	假设条件		如果(假如、假使、假若、假设、假定、倘若、倘使、若是、若、要(不)是、若非、万一、的话……)	那、那么、则……	再、就、便、也……
	有条件	充足条件	只要(只需、一旦……)		就、都、便、总……
		必要条件	只有(唯有、除非……)	否则、不然、要不然	才

表16.2 英语中表示条件关系的连词

(Quirk et al 1985)

连接关系类型		句首	句中	连接副词	连接语
条件	假设条件	if, assuming (that), given (that), supposing (that), if only	if, given (that), supposing (that)		if so
	强加条件	unless, as long as, so long as, providing/provided (that)	unless, as long as, so long as, on condition that, providing/provided (that)		if not, otherwise

我们以每一个表示条件关系的英语词汇在汉译英语料《围城》中进行检索,去掉干扰因素,共获得539个句对(如表16.3)。

表16.3 汉英翻译中条件关系句子对应情况

(以《围城》英译为例)

翻译方向	关系类型	文本名称	总句数	汉英对应		连接标记 ↓ 条件连词		零标记 ↓ 条件连词	
				句数	比例(%)	句数	比例(%)	句数	比例(%)
汉译英	条件关系	《围城》	539	193	35.81	131	24.30	215	39.89

以if为例,其中193个句子是由汉语原文中的条件连词对译而来。如:

(32) **假使**这样就会爱上一个人,**那么**,爱情容易得使自己不相信,容易得使自己不心服了。

If I could fall in love with someone that easily, ***then*** I can't either

believe in or submit to love.

(33) 你**若**补写信来慰问，好比病后一帖补药，**还**是欢迎的。

If you wrote a letter inquiring after my health it will still be as welcome as an extra dose of medicine after sickness.

(34) 鸿渐气得把手杖残暴地打道旁的树不如趁早死了心罢，给一个未成年的女孩子甩了，**那**多丢脸。

So vexed by this fact, he struck his cane violently against a roadside tree and decided he'd better quash all hope from the very start. What a disgrace it would be ***if*** he were to be jilted by a teen-age girl.

(35) 内地买不到西药，各位**万一**生起病来，那时候**才**知道我李梅亭的功劳呢！

You can't get Western medicine in the interior. ***If*** by chance someone should get sick, ***then*** he'll appreciate what I, Li Mei-t'ing, have done.

131 个句子是根据汉语中表示条件关系的标记词，如连接副词，在译文中以条件连词加以明示。如：

(36) 所以病人怕死，**就**得请大夫，吃药。

So ***if*** a sick man is afraid of death, he'll call a doctor and take medicine.

(37) 有了下半箱的药，中国人全病死了，李先生**还**可以活着。

And with the medicine in the bottom half, ***if*** the Chinese all died of disease, Mr. Li could still stay alive.

(38) 你把事实聚拢来**就**明白了。

If you put all the facts together, you'll understand.

(39) 你自己不好意思，托旁人转问一下**也**行。

If you feel uncomfortable about asking him yourself, you could have someone else ask him for you.

(40) 我看见自己的丈夫给人家笑骂，**倒**实在受不住，觉得我的脸都剥光了。

Well, ***if*** I ever saw my husband being ridiculed and abused, I certainly wouldn't he able to put up with it. I'd feel as if I had been stripped bare.

而其余的 215 个句子，汉语原文中没有连接标记，译文进行了连词添加。如：

(41) 三闾大学的电报，家里还没知道，报告了父亲母亲，准使他们高兴，他们高兴头上也许心气宽和，不会细密地追究盘问。

His family still didn't know about the telegram from San Lü University. What ***if*** he announced this to his parents? They would certainly be pleased, and in their joy they might be more magnanimous and not probe and question him very much.

(42) 跟着他走，准闯祸。

If you go with him, you're sure to meet with disaster.

(43) 有话斯斯文文地说。

If you have something to say, say it like a gentleman.

(44) 今天不早睡，明天上岸的时候人萎靡没有精神，难看死了。

If I don't go to bed early, I'll be all tired out and will look a wreck when I go ashore tomorrow.

(45) 一天清早，专差送信来，是赵辛楣写的，说昨天到点金银行相访未晤，今天下午四时后有暇请来舍一谈，要事面告。

One day early in the morning a special messenger came with a letter from Chao Hsin-mei. The letter said that the day before Chao Hsin-mei had gone to the Golden Touch Bank to see him but had not found him there. ***If*** he were free this afternoon, could he please drop by after four for a chat? Hsin-mei had something important to tell him.

(46) 咱们不打不成相识，以后相处的日子正长，要好好地交个朋友。

We wouldn't have known each other ***if*** we hadn't had a fight. We'll be together for a long time later on, so we should try to be good friends.

(47) 你不信，问他自己。

If you don't believe me, ask him yourself.

(48) 阿福威风百倍道："你有种出来。"

"Come out here, ***if*** you have the guts!" Shouted Ah Fu with exaggerated authority.

(49) 让她去造谣言得了，反正没有人会相信，我就不相信。

Let her spread rumors ***if*** she wants. No one's going to believe them anyway. I for one don't believe them.

(50) 譬如一串葡萄到手，一种人挑最好的先吃，另一种人把最好的留在最后吃。

If given a bunch of grapes, for example, one kind will eat the best ones

first, while the other kind will leave the best for last.

(51) 今天灌醉了你,对不住孙小姐的。

If I got you drunk today, I wouldn't be doing right by Miss Sun.

(52) 鸿渐道:"累了快去睡,我不讲了。"

"***If*** you're tired," he said, "then go to bed. I won't talk anymore."

(53) 你要吵,随你去吵。

If you want to quarrel, then go ahead.

(54) 你愿意挨饿,活该,跟我不相干。

If you want to go hungry, it serves you right. It makes no difference to me.

(55) 小心别讨了你那位好朋友的厌,一脚踢你出来,那时候又回上海,看你有什么脸见人。

Just be careful you don't annoy this friend of yours. ***If*** he kicks you out, and you come back to Shanghai, I'd just like to see if you can face anyone then.

由以上对译例句可以发现汉语中包含条件关系的句子中并不一定要有显性连接词,而是依靠连接标记、语序、语调等加以传达,而在译成英语时往往会在语言表层加以体现。

反过来,我们以英文文学文本《老人与海》及其汉译为例,考察 *if* 一词可以发现,原文中包含 if 的条件句共有 67 句,在汉语译文中有 59 个都以汉语表示条件的连接词对译出,只有 8 句隐去了显性的连接词。但是,我们仔细观察一下以连词加以对译的句子,就会发现许多都可加以修改,从而更加符合汉语的表达。如:

(56) "***If*** the others heard me talking out loud they would think that I am crazy," he said aloud.

"**要是**别人听到我在自言自语,会当我发疯了,"他说出声来。

(57) "But I will say ten Our Fathers and ten Hail Marys that I should catch this fish, and I promise to make a pilgrimage to the Virgin de Cobre ***if*** I catch him."

"但是我愿意念十遍《天主经》和十遍《圣母经》,使我能逮住这条鱼,我还许下心愿,如果逮住了它,一定去朝拜科布莱的圣母。"

(58) But ***if*** I eat him fresh enough he won't be bad.

但是,**如果**趁它新鲜的时候吃,味道不会差。

(59) It was these sharks that would cut the turtles' legs and flippers off when the turtles were asleep on the surface, and they would hit a man in the water, ***if*** they were hungry, even if the man had no smell of fish blood nor of fish slime on him.

就是这些鲨鱼,会趁海龟在水面上睡觉的时候咬掉它们的脚和鳍状肢,**如果**碰到饥饿的时候,也会在水里袭击人,即使这人身上并没有鱼血或黏液的腥味。

(60) ***If*** you love him, it is not a sin to kill him.

如果你爱它,杀死它就不是罪过。

将以上汉译句子中的"要是"、"如果"去掉,基本上不会影响句意的表达,反而更为简洁。

16.3.2 结构的对译检索

英汉语都有一些特殊的句型和结构,在翻译时如何处理,不是教科书所能讲清楚的。需要许多的例句给学生以感性认识。双语对应语料库可以做这方面的补充。例如对汉语"把"字结构的翻译。我们通过对双语平行语料库的考察,发现把字结构通常有三种形式(王克非 2003)。如何将其翻译成英语,有哪些译法?若以双语库作一对译检索,可以增加学生对该结构翻译的感性认识。例如以下对译:

(61) 很娇媚地**把**头**一扭**,她又吃吃地笑着。

She *tilted* her head coquettishly and giggled.

(62) 杜竹斋**把**舌头**一伸**,嘻嘻地笑了。

Tu -Chu-chaff *stuck out* his tongue, withdrew it, and grinned.

(63) 我们**把**恐怖分子**打得抱头鼠窜**。

We *have* the terrorists *on the run*.

(64) 都穿上大褂,谁短撅撅的进来**把**谁**踢出去**。

You're all to wear long gowns, anyone who shows up in a short jacket *gets kicked out*.

(65) 他估计可以趴着挤进去时,就**把**头先**伸进去**……

He judged he might squeeze through on his knees; so he *put* his head *through*...

(66) 和大家一齐坐下,大家**把**对刘四的不满意都**挪到**他身上来。

But as soon as he sat down the others *transferred* their resentment

toward Fourth Master Liu *to* him.

(67) 甚至当他还是个年轻的孩子时，就能**把**表**拆开**又**装上**。

Even when he was still a young boy, he could *take* a watch *apart* and *put* it *together* again.

(68) 他**把**一支香烟**放在**嘴里，一半烟丝就掉在舌上。

He *put* a cigarette *in* his mouth. Half the tobacco promptly fell out on to his tongue.

(69) 他**把**自己**形容成为**一个软心肠的人。

He *describes* himself as an old softie.

(70) 经济调研**把**这一点**归因于**劳动力市场日益灵活。

Economic research *puts* this *down to* increased flexibility in the labour market.

(71) 他们喜欢**把**自己**打扮成**一无所有的愣头小子。

They *prefer an image of* penniless tough guys.

(72) 他愿意快快**把**这一天**过去**，不再受这个罪。

So he wished the day would *end* quickly to end his discomfort.

(73) **把**鞭子**递给**我。

Hand me that switch.

(74) **把**上衣的纽扣**解开**！

Unbutton your jacket!

由上述例句，总结汉语把字结构的译法为：

第一，选择适当的英语使动动词，包括用 make/have/get 等相关搭配(如(61)～(64))；

第二，仿汉语形式：英语动词＋宾语＋补语，或动词＋宾语＋介词短语(如(65)～(70))；

第三，其他变通方式(如(71)～(74))。

通过这样大量的对译检索，学生可以对"把"字结构及其翻译又直接的感受，教师再加以适当的方法上的归纳和解释，就能取得较好的教学效果。

还可以进行近似结构的对译检索。关于这一点，台湾大学高照明自建了一个英汉平行语料库，尝试进行近义句的翻译检索。学生可以输入中文关键词，从平行语料库中搜寻包含该关键词的英文句子及其译文，或反之，输入英文词语而找寻对应的中文译文。以"他心情很低落"为例，从语料库中可寻得如下相近的中英文短语、句子(高照明 2004)：

心情郁闷、烦躁，情绪低落—get the hump
这几天他似乎情绪十分低落。—His spirits seem to be drooping these days.
那消息使士兵们士气低落。—The news chilled the morale of the soldiers.
士气高[消沉，低落]—Morale is high [low, falling].
我今天情绪低落[高昂]。—I feel down [up] today.
他的心情变化不定。—His moods change quickly.
忧虑使他心情抑郁。—Cares oppressed his spirits.
他心情沮丧地回家去。—He went home, dejected in heart.
他现在脾气[心情]不好。—He's in an ill temper [humor].
他心情不好，拿妻子出气。—He vented his ill humor on his wife.

这些显然也是一般词典所不能提供的帮助。

16.3.3 一本多译的比照和参考

双语对应语料库还往往能提供一本多译的丰富译例。限于篇幅，仅举一两例。如下是王宗炎先生曾谈过的例句：

> The cause of the Republican Party's defeat in the election is perhaps best summed up by a Scottish writer John Arbuthnot, who said, "*All political parties die at last of swallowing their own lies.*"

其中最后一句斜体部分的翻译有多种：

一切政党最后都因吞下了自己的谎言而死。
所有政党最后不免死亡，都是由于谎言梗塞了喉咙。
一切政党最后混不下去，原因不外乎食言太多。
虚言假语堵住了嗓子，这是所有政党最后失败的根由。
一切政党都因为说话不算数，终于垮台。
所有政党全是这样：大话说尽之日，即生命告终之时。
一切政党的结局都是，骗局拆穿，政权丧失。
一切政党到头来都是，空头支票吃不开呜呼哀哉。

在翻译教学中，可以让学生自己试译，并给出所有这些译例供学生评论。当然，这里还可提出其他译法，例如：“所有政党最终都因为自食谎言而灭亡。”这样的

练习，学生会有兴趣，并从中得到良好的翻译训练。

因为双语对应语料库解决了大规模语料存储的问题，对于翻译工作者或翻译教师来说，最为便利的莫过于将所收集的多种译本存储于其中，并可随时调出加以对比研究。名著重译或复译是非常多的，一般人很难收集齐全，如《简爱》有近 20 个译本，《傲慢与偏见》的译本也有十六、七种之多，《红与黑》译本据说有人收集了 20 个以上，莎士比亚戏剧中《汉姆雷特》的汉译本最多，达到 30 种。还有诗歌的译本，无论是英诗汉译还是汉诗英译，都常有多种译本，李白《静夜思》就有十多种译文。将多种译本收集到平行语料库或翻译语料库里，就会大大方便翻译的研究者和教师，对学生进行学习、观摩和比较非常有益。孤立地看待一个译本，往往不能看出什么，多个译本综合到一起看，就可能有新的发现，对翻译对比和翻译教学有所帮助。

16.4 小结

以上从三个方面的大量例证，阐释了双语对应语料库在翻译教学上的价值，表明它的确有助于提高学生理解原文并用更流畅的目标语进行翻译的能力。一方面可以帮助学生做外译汉练习，参照对应语料库查找相应的译文，锻炼和提高自己的翻译表达能力。另一方面，学生做汉译外时，会更多地利用双语对应语料库，特别是其逆查的方法，学生可以参考和借鉴已有译法，有时还有多个翻译选择供参考，使术语和短语的翻译更加准确，译文也更加地道。此外，双语对应语料库使得自动翻译的研究前景变得明朗，反过来也会对今后的翻译教学有更大的助益。

第 17 章

对应语料库翻译教学平台的应用

本章通过语料呈现实验探讨双语对应语料库翻译教学平台的应用效果。实验显示，学生在观察语料后能够归纳和总结出有意义的翻译技巧，并能据此评估或反思自己的翻译行为。实验表明，在翻译课堂教学中使用语料资源有助于自主学习和发现式翻译教学环境的创建，也有助于学习者形成稳定的翻译技巧。

17.1 引言

近 20 年来，语料库的创建和应用取得了长足发展。对应语料库大都在 20 世纪 90 年代开始创建，起步较晚，但已展现出广阔的应用前景。在欧美，这类语料库有十多个，涉及近 20 个语种，如 Hansard（法-英对应语料库）和 ENPC（英语-挪威语对应语料库）（Véronis 2000：14-15）；在中国，有中科院计算所的大规模汉英对应语料库、北京大学的“BABLE 汉英对应语料库”，以及北京外国语大学的“汉英对应语料库”（该库现有可检索语料 2 000 万字/词以上，见王克非 2004a），后者是本项研究使用的主要工具。

对应语料库的创建有两个主要目的：一是用于语言与翻译研究，二是用于外语教学。用于前者的研究成果丰富，涉及语言对比、双语词典编纂和翻译研究（Laviosa 1997；Baker 1999；Serpellet 2000；Hunston 2002）。用于后者尤其是用于翻译教学的研究也已取得不少成果，如 Zanettin 1998、Pearson（2003：15-24）和 Bernardini（2004b：97-111），但大都研究平行语料为翻译训练提供的可能的资源和手段，还没有研究探讨大型对应语料库在课堂教学中的实际应用问题。

17.2 研究内容

有关运用语料库提高翻译效率和质量的研究还没有全面展开，而将对应语料库应用于翻译教学的研究则是刚刚开始（Bowker 2003；Bernardini 2004b）。从理论上看，语料库用于教学有利于自主学习环境的创建（Bernardini2004a：22；秦洪

武、王克非 2007)。自主性学习主要表现在两个方面:研究性学习和发现式学习(learning as discovery)。前者假定学习者和教师研究兴趣相投、研究能力相当;后者则鼓励学习者自行调节兴趣点,并给他们提供机会来提高自己观察和处理问题的能力,使他们对两种语言的特征和差异有敏锐的觉识。本章认为,研究和发现虽不矛盾,但在以技能培养为核心的翻译教学中,不宜过分强调学生的研究能力,发现式学习这一提法更合适一些。

近年来,国内也有研究关注语料库在翻译教学中的应用问题。有些研究探讨了语料库在翻译教学中的用途(郭红 2004;于连江 2004;王克非 2004b),但还没有研究系统地探讨对应语料在课堂教学中的应用方式和应用效果,也没有研究关注学生对于语料使用的态度。鉴于此,我们以"汉英对应语料库"为翻译教学平台,探讨在现有技术条件下翻译语料用于翻译教学可能产生的积极效果。

17.3 研究设计及步骤

实验使用的翻译语料全部取自北京外国语大学"汉英对应语料库"。语料呈现实验分为两步。第一步要求受试观察真实翻译语料;第二步让受试将自认为有意义的发现以书面形式表达出来,鼓励同学之间(一般为同桌)通过商讨共同完成观察报告。

17.3.1 参与者

参加本次调查的是某大学外语学院英语专业本科四年级学生(三个班,136人)和一年级研究生(35 人)。这些学生都在学习翻译,此前没有接触过语料库。本研究主要观察学生对语料呈现的整体反应,因此,个体和性别差异未列入考察范围。但是,鉴于语言水平可能影响到语料库的使用,在整理数据时将研究生和本科列为对比分析的对象。

实验分三次进行。第一次的受试是本科四年级三个自然班的学生,呈现的汉英对译语料涉及汉语动词与英语介词 *in/at* 间的对译关系(呈现给 1、5 班),以及汉语动词与 *over/with* 之间的互译①关系(呈现给 2 班)。为利于把握受试对语料的

---

① 如:
隔壁绍兴戏唱完了,你就打鼾,好厉害
After the Shaohsing opera was [over], you started snoring your head off
韩孟翔狡猾地笑着回答,把手杖一挥,就沿着那水门汀向南走,却故意放慢了脚步
he replied [with] a sly grin, then turned away down the street, swinging his stick, though he was careful not to walk too fast

反应，实验使用的对应语料由教师课前筛选。在呈现语料时，教师不对学生做任何暗示，以保证学生自主做出判断。第二次的受试是本科四年级学生，第三次的受试是一年级研究生，这两次实验使用的翻译语料都是汉语句群的英译①。以上三次语料呈现方式相同。具体步骤为：

第一，语料呈现——刺激：将语言学习者的注意力集中于关键词或结构的翻译对应项。语料由教师挑选（也可以应学生要求或建议学生自行检索）。

第二，学生浏览语料：学生通过考察语料建立自认为有效的形式-意义关联。

第三，提示学生记录他们的心得和他们认为有意义的发现。

第四，要求学生回答一个开放性问题：观察语料后你有什么心得或发现？

让受试回答上面的开放性问题有两个目的：一是考察学生对语料呈现的态度；二是观察学生在浏览语料时关注了哪些对翻译教学有意义的问题。

17.3.2 数据收集

本项研究呈现的语料量为 1 500～1 700 个汉英字词（汉语字与英语词比率为 1.6∶1），学生可在十分钟左右浏览完毕。阅读完毕，教师要求受试回答一个开放性问题：观察语料后你有什么心得或发现？教师建议报告由两人或多人共同完成，长短不限。学生的回答以书面报告形式提交给老师，供下一步进行归类和分析。

17.3.3 数据分类

鉴于学生回答的问题是开放式的，我们采用吴一安（2005）使用的主题一致分析方法对学生回答的内容进行分类，分类后再整理出受试关注的内容细节。这些内容以提及频数的形式呈现出来，形成附录中的六个统计表。

17.3.3.1 受试对语料呈现的态度

三次实验的对象都是英语专业的学习者，他们正在学习汉英翻译。报告显示学生对于汉英翻译时的处理技巧比较敏感，而且也表达了自己对于语料呈现的态度。如：

---

① 如：

◆〖4∶1〗

<s>是一条快船。</s>

<s>单趟直放，不过半天多，就到了；</s>

<s>就是颠得厉害。</s>

<s>骨头痛。</s>

<s>It's a fast boat and didn't stop on the way, and we did the journey in half a day, but it did roll terribly, and made my joints ache. </s>

第一，英语介词很有用处，有时比一个句子表达得还要清楚。

第二，看似挺难的汉语句子，用英语的介词短语简简单单就译出来了，而且带有原文的韵味，其乐无穷。

第三，平时翻译时意识不到用英语介词使译文简洁。

第四，翻译不仅要靠大的英语词汇量，还要学会灵活使用我们已经掌握了的词汇和语法。

第五，看这些汉英对译很有趣，也感觉很有用。

可以看出，受试在观察语料后不仅能够意识到语言运用问题，还能够反思自己的语言能力和翻译能力，有的甚至指出需要调整自己的语言学习方式。总的说来，学生对翻译语料呈现的反应是正面的。

17.3.3.2 报告内容的分类

我们把受试报告的内容分成四类：技巧觉识、功能认识、语言差异和评估（对翻译手段的评价）。

第一，技巧觉识：受试注意动词与介词之间的转换关系，也注意到汉语句群在英语译文中的组合方式。这说明受试具有从对应语料中觉识到翻译技巧的能力。

第二，功能认识：受试注意到一些英语的介词能够引入时间、状态、年龄、同时性等背景信息，还注意到英语连词、从句在语言组织上的独特作用。这说明受试对技巧的觉识与他们对词汇和语法功能的认识是相通的。

第三，语言差异：有些受试注意到英汉两种语言的差异。

第四，评估：学生注意到了某些翻译处理方式产生的文体效果和语言运用的灵活性，说明学生对翻译技巧有评价能力。

以上分类说明，在呈现汉英翻译语料后，受试对于翻译技巧、词汇-语法功能、语言差异比较敏感，同时也注意到了翻译技巧的运用效果。这些都表明，受试能够有效地观察和评价翻译语料，翻译语料呈现的作用是积极的。

17.4 结果与分析

17.4.1 第一次呈现的语料

第一次语料呈现后，收到的 64 份报告中有 41 份（占 64%）提及汉语动词和英语介词短语在翻译中的转换问题（见表 17.1 至表 17.3）。表中显示，有关技巧觉识方面的内容出现频率最高，也最集中。统计还显示，受试对于介词的功能也做了较为全面的归纳，但多停留于对共现词项的描述，不如对技巧的觉识那么集中。另外，表 17.1 至表 17.3 显示，受试陈述个人对译文的评价时主要涉及文体效果（19

次)和介词使用的灵活性(六次)。

统计还显示,受试对于不同类型的介词有不同的认识,这一点主要体现在对*over*、*with* 这两个介词的表达功能上。表 17.3 显示受试注意到介词的功能是引入背景信息,说明部分受试在观察翻译时已发现汉英翻译时动词与介词转换的内在依据,同时也注意到英语在表达背景信息时可以使用与汉语不同的语言手段。上述内容表明多数受试能从呈现的语料中归纳出有意义的翻译内容,并能给予比较合理的解释。

当然,受试报告中对于译文的评估有失偏颇。如表 17.1 有两份报告、表 17.2 有一份报告认为,汉语生动、英语单调。做出这一概括的主要依据是:呈现语料中的汉语动词群都转换成了英语译文中的"介词+名物化"短语形式。但这是一种误解,教师需要对此做出分析和说明。

面对成批出现的语料,只有少数受试明确而具体地提及英汉语言之间差异,说明大多数受试还不能在更抽象的层面上思考语言差异问题。

17.4.2 第二次呈现的语料

表 17.1、表 17.2、表 17.3 与表 17.4、表 17.5 之间存在有趣的差异。表 17.1、表 17.2、表 17.3 中的主要内容是词的运用和功能,而表 17.4、表 17.5 则主要涉及句子和句际关系。显然,观察语料的类型不同,受试关注的内容也就不同。例如,呈现的多为句对应语料时,受试关注的重点在词和词性(见表 17.1、表 17.2、表 17.3);当呈现内容为句群对应语料时,受试则将关注的重点调整到衔接手段的使用上(见表 17.4、表 17.5)。总之,受试能够根据呈现的材料来调整自己的注意焦点。但表 17.4、表 17.5 也显示,个别受试的归纳和概括还有些问题。例如,有受试认为英语译文比汉语原文逻辑性更强。这一判断只是依据连接词使用的数量,但逻辑性的强弱不是凭连词使用的数量来判断的。

本科生受试的报告显示出他们对语料感兴趣,也能捕捉到潜藏在译文后面的翻译技巧。这至少可以说明,本科生能够通过观察翻译语料、归纳并通过讨论来学习翻译技巧。这些技能和心得在初期需要教师有针对性地提供语料,但归纳和总结都是学生自主完成的。从这个意义上说,翻译语料的呈现可以有效地帮助翻译学习者获得翻译技能、提高翻译水平。

此外,语料呈现还有助于学生调整学习方法,注意自身语言运用能力的培养。如有受试在报告中写道:

翻译不仅要靠大的单词量,还要学会灵活使用我们已经掌握了的词汇。

有个别受试甚至还觉察到自己的语言直觉和实际运用之间有距离。如:

我学了以前不知道的东西。发现一个简单的介词有不同的意思和功能。一个

句子可以有不同的表达方式。这些译文处理技巧很精彩。看来我还要学更多的东西。这些技巧能够帮助我们译出好的译文。

分析真实文本可使学生在词汇和语法结构方面增长见识，这有利于学习者承担真实的翻译任务。观察语料不仅能够激发学习动机，还有助于学生适时调整翻译学习方式，提高对翻译技巧的认识和应用水平。一般说来，通过考察一个词或者结构的多个翻译例证，个别性的因素就会被规律性替代；有动因的翻译策略累积起来就可使学习者感知到一般的型式并形成有意义的概括。

17.4.3 第三次呈现的语料

第三次呈现的语料与第二次相同，但受试不同。这一次的受试是一年级学习翻译理论与实践的研究生。表 17.6 显示，受试对于翻译技巧细节的捕捉比本科生(表 17.5、表 17.4)受试更细腻。而且，他们对于语言事实抽象能力更强。例如，受试在报告中使用衔接手段、重构、从句这类术语来表述问题，而少用生动、地道等含混的字眼来表达自己的看法。

17.4.4 报告反映出的问题

呈现语料可能引发不当的归纳和概括，原因主要来自两个方面：一是语料呈现的数量有限，无法涵盖与某一语言项目相关的所有情形；二是限于语言能力和分析能力，受试还无法深入思考语言和翻译问题。出现这些问题是正常的，但教师应该适时加以引导，鼓励学生尽可能全面地观察翻译和语言现象。

本次实验使用的语料是教师检索后筛选出来的，还不是由学生根据自己的需求自行检索。因此，这还不是真正意义上的自主性学习。要想充分挖掘学生的学习动机和潜力，就应鼓励学生按照自己的需求去检索、获得和研究真实语言素材，并提出自己的看法。这就为学生作为"研究者"去开辟新的研究领域带来了实实在在的希望，而不只是重复或者评估别人的研究(Leech 1997:10)。

目前在课堂上只是以电子文本或者检索软件的 KWIC 形式呈现翻译语料，这样的呈现方式完全由教师控制，过于机械、单调。鉴于此，创建方便、迅速的语料检索和应用环境，并实现学习者按照个人的需求检索、考察、探讨和使用语料，是对应语料库有效应用于翻译教学的关键。

17.5 小结

Bernardini(2004a:31-2)指出，语料库和语料分析工具似乎为发现式学习活动提供了强有力的辅助工具。规模较大的对应语料库能为翻译教学提供丰富多样的

翻译素材，为灵活运用翻译技巧提供重要的参照资源，也为建立良性的课堂互动提供重要的物质环境。这些都有利于激发学生的学习兴趣，促使学生将注意力集中在与技巧密切相关的形式-意义关联上。发现式学习本来不是基于语料库的，但语料库却能够有力地促进自主学习环境的建立。

学习者的观察结果可能是局部或者经验式的概括，但这一观察结果可以通过新的证据来检验、修改和丰富。从外语学习和翻译技巧觉识上看，学习者大量观察双语对应语料后可能对搭配、语义偏向和语义韵等现象更加敏感；翻译技巧和策略可以通过学生之间和师生之间的互动性讨论逐步形成，这有助于学生形成稳定、持久和灵活的翻译策略。传统上，教师在翻译教学过程中起主导作用，面对教材中数量有限且少有例外的零散范例，学生往往会盲目接受教师所陈述的规定性很强的翻译技巧，而翻译质量评估则完全依赖教师的直觉判断。这些都无助于培养学生灵活运用翻译技巧的能力。使用语料库可以弥补这一方面的不足，它允许学生（根据教师的建议）自主且有针对性地选择语料并藉之做出概括或解释，充分发挥其自主捕捉翻译技巧和自主构建翻译策略的能力。从这个意义上说，对应语料库的应用有助于提升翻译教学的效果。

17.6 附录:受试报告的分类和统计

17.6.1 第一次呈现的语料:汉英翻译时介词短语的使用

表 17.1 汉语中的动词表达转换为英语 in、at 引导的介词短语

(本科四年级 1 班,共收取报告 24 份)

报告内容分类	技巧		功能					语言差异	评估			
	动词与介词转换	有些需要更大语境才明白	at 表同时性	at 表时间点,in 表时间段	at 引出一个行为	at/in 可以表达年龄/状况	in 可表方位、状态	英语多介词,汉语多动词	反思自己的翻译	英语介词用法灵活	汉语生动,英语单调	文体效果
频次	18	1	1	7	1	4	11	1	1	2	2	3

表 17.2 汉语中的动词表达转换为英语 in、at 引导的介词短语

(本科四年级 5 班,共收取报告 21 份)

报告内容分类	技巧	功能			语言差异	评估		
	动词与介词转换	at 表时间点,in 表时间段	at/in 可表年龄/状况	in 可表方位、状态	英语多介词,汉语多动词	文体效果	英语介词用法灵活	汉语生动,英语单调
频次	10	3	5	6	1	8	2	1

表 17.3 汉语中的动词表达转换为英语 over、with 引导的介词短语

(本科四年级 2 班,共收取报告 19 份)

报告内容分类	技巧	功能			评估	
	动词与介词转换	over 表时间	over/with 可以表达状态	介词引人背景信息	文体效果	用法灵活
频次	13	2	1	7	13	4

17.6.2 第二次呈现的语料:汉语句群的英译

表 17.4 本科四年级 5 班

(共收取报告 20 份)

报告内容分类	技巧				语言差异		评估	
	汉语短句译为英语小句	英译文添加了连词	用介词引入从属性行为	用分词短语表达从属性行为	英语句子比汉语长	汉语省略多	英译文逻辑性更强	文体效果
频次	11	15	6	8	2		1	8

表 17.5 本科四年级 2 班

(共收取报告 17 份)

报告内容分类	技巧							语言差异			评估
	汉语短句译为英语小句	英译文添加了连词	用介词引入从属性行为	添加/改变主语	语序调整	重新组合	反思自己的翻译	英语句子比汉语长	汉语省略多	汉英比英汉使用技巧更多	英语译文逻辑性更强
频次	6	12	2	2	5	4	1	2	1	1	1

17.6.3 第三次呈现的语料:汉语句群的英译

表 17.6 研究生一年级

(共收取报告 22 份)

报告内容分类	技巧										语言差异		评估
	遣词	汉语短句译为英语小句	英译文添加了连词	译文使用副词表达原文的行为	汉语重复而译文使用代词	使用从句	使用了原文没有的衔接手段	调整语序/重构	背景与前景	添加/改变主语	英语句子比汉语长	汉语少用衔接手段	介词使英译文简洁
频次	1	11	11	5	3	6	4	4	1	4	6	4	9

第 18 章

从语料库看翻译搭配及其教学

搭配是重要的语言组织原则，这一点在语料库语言学中得到进一步证实。汉译英时，一方面由于受汉语原文中搭配模式的影响，另一方面由于英语不是母语，会导致英语译文中的搭配不当，进而影响译文质量。本章应用汉英对应语料库，从英译汉和汉译英两个方向分析了“克服”一词在汉语中的各种搭配及其在英语中的对应形式，阐述了对应语料库在汉译英中丰富学生搭配知识的作用。此外，作者还阐述了目标语英语的单语母语语料库如 BNC 在汉译英搭配教学中的作用。

18.1 引言

搭配指的是语言横组合轴上词与词的组合。Firth 首先发现词语之间的横组合关系对于揭示意义有重要作用。Firth 的名言就是“要了解一个词就须了解其结伴词”(Firth 1957:12)，如 *ass* 的部分意义存在于其习惯性同伴词 *silly*, *obstinate*, *stupid*, *awful* 中(同上:195)。进而，Firth 提出了“搭配”概念，即词语的习惯性结伴(同上:99)，被后人誉为“搭配之父”。现代语料库语言学通过对词汇的考察表明，搭配是重要的语言组织原则，恰当的搭配已成为区别母语与非母语的重要语言标志(Cowie 1994;Channell 1980)。搭配无疑已是语料库语言学中最为活跃的研究领域，处于它的中心地位(杨惠中等 2002:82)。

但每一种语言的词汇往往都有自己的独特的搭配模式，在进行跨语言翻译时，人们往往不能根据其组成成分直接进行翻译。所以，搭配的翻译是翻译过程中的一个大问题(Baker 1992;Malmkjær 2005;Bahumaid 2006)。Bahumaid(2006)指出了对搭配进行翻译可能遇到的两大类问题：语内问题和语际问题。语内问题主要包括：1)搭配尚无严格的定义，因此搭配与非搭配间的界限不清，例如，我们没有明确的参照标准判断以下组合是搭配还不是搭配：*a good/strong*/high chance**, *a good*/strong/high probability* 和 *a good/strong/high likelihood**；2)两个或多个同义词在一语境中可能重叠，而在另一语境中可能相互排斥，如 *remote* 和 *distant* 都可与 *past* 搭配，但 *remote* 只能与 *village* 搭配，*distant* 只能与 *memory* 搭配；3)有些

搭配是一种特定语域所独有的，如 *dull highlights* 主要用于气象学和摄影；4)新的搭配不断出现，因此译者须不断学习新的搭配；5)搭配存在一个度的问题，对某些人来说一种搭配是可接受的，而对另外的人则不然。语际问题主要包括：1)一种语言与另一种语言中的词汇搭配能力不同；2)原语中的自由短语在目的语中可能是搭配，或者相反，如英语：*to overlook something*，阿拉伯语：غض الطرف عن شيء①；3)词语在原语和目的语中的搭配范围不同，如英语：*catch a fish/a cold/a train/fire*；阿拉伯语：يصطاد سمكة يصاب بنزلة برد يلحق بالقطار تشتعل النار فيه②此外，我们认为，搭配具有内隐性，不能直接观察，往往需借助大型的语料库才能发现。在汉译英教学中，一方面由于原文搭配模式的影响，另一方面又由于英语不是母语，学生的搭配知识不够丰富，所以学生译文中出现较多的非典型，甚至是目的语不可接受的搭配模式，导致译文中搭配不当，进而影响译文质量。

18.2　汉译英的搭配问题

汉英两种语言搭配模式差异较大，汉译英时往往不能直接翻译，否则就会出现翻译中的搭配错误。根据我们的观察，汉译英中常见的搭配错误主要包括：主谓搭配错误；动宾搭配错误；定中搭配错误。

18.2.1　主谓搭配错误

英汉语存在语言类型学上的差异：汉语是话题突出语言（topic-prominent language），英语是主语突出语言（subject-prominent language）（Li & Thompson 1981；曹逢甫 1979）。实际上，汉语中的主语有许多情况下是话题，而不是主语，因此和谓语的关系较为松散，语义、语法限制较少。但英语中话题句较少，主语和谓语的关系紧密，语义、语法限制较为严格。例如：

（1）我们国家的**建设事业**不仅仅**需要**的是工程师和科学家，同样需要各行各业的能工巧匠。

The **building cause** of our country **needs** not only engineers and scientists, but talents of various kinds.

Our **country** in her development **needs** not only engineers and scientists,

---

① 在英语中 *to overlook something* 是自由组合，但在阿拉伯语中是搭配。

② 英语中的同一个 *catch* 到阿拉伯语中换成了三个不同的动词。

but talents of various kinds.

(2) 这个**战略思想**现在已经**有了**实施的机遇。

The **strategy** now **has** an opportunity to implement.

We now **have** an opportunity to implement the strategy.

上述两句在英语中的"need"和"has"都需要有灵主语,而汉语则无此要求。因此,不懂得这种主谓搭配限制,就会在汉译英时出错。

18.2.2 动宾搭配错误

(1) 我想我们不能**同意**这样的**要求**。

I'm afraid it is impossible for us to **agree with** such **requests**.

I'm afraid it is impossible for us to **grant** such **requests**.

(2) 有必要**教给**学生一些美国历史的**知识**。

It is necessary to **teach** students a little **knowledge** of American history.

It is necessary to **impart** students a little **knowledge** of American history.

(3) 帝国主义和国民党反动派的长期统治,**造成了**社会经济的**不正常状态**,**造成了**广大的**失业群**。

Long years of imperialist and reactionary Kuomintang rule **left** our economy in a chaotic **state** and **caused** widespread **unemployment**.

例(1)和例(2)中的"agree with"和"teach"的宾语应为人,这反映了汉英搭配模式的因循性差异。例(3)中"cause"一般表示造成某种结果,不能表示状态,所以"造成状态"就需换为"left"。

18.2.3 定中搭配错误

(1) 我是超市的**老顾客**,所以也不抱怨什么。

As an **old customer** of the supermarket, I did not make a complaint.

As a **regular customer** of the supermarket, I did not make a complaint.

(2) 我们要**广泛普及**科学文化知识。

We need to **extensively popularize** science and culture.

We need to **popularize** science and culture.

例(1)中的"old"在英语中并无频繁的意义,因此和汉语原文不对应。例(2)中的"popularize"本身就已包含"广泛地"的意义,因此在英语中无需再加"extensively"。

由此可见，汉英两种语言在搭配模式上存在较大差异。汉译英中，搭配的翻译是一个大问题。在这方面，我们认为，语料库是汉译英中帮助学生丰富搭配知识的重要手段。

18.3 基于语料库的汉译英中的搭配教学

对于汉译英中的搭配教学，我们认为两种语料库有较大作用：汉英对应语料库和目的语为英语的单语母语参考语料库。

首先，我们看一下汉英对应语料库对汉译英中搭配教学的作用。我们以“克服”一词的英译为例进行说明。汉译英中，汉语“克服”一词最常用的英语对应词是“overcome”。但通过对汉英对应语料库(CEPC)的检索，我们发现“overcome”在英语中与“克服”在汉语中的搭配模式有较大差异。汉语中“克服”的许多搭配模式在英语中无法与“overcome”搭配。汉英对应语料库中“克服”的搭配模式见表18.1。

表 18.1 “克服”的中文搭配词

	汉英对应语料库中“克服”的中文搭配词
克服	现象，困难，矛盾，阻力，观念，官僚主义，危险，不平衡，风气，逆转，思想，旧毒，污浊，波折，回头路，意见，缺点，贪污，行贿，盗窃，东西，错误

我们发现，搭配词中除了“困难”，“阻力”，“缺点”外，其他汉语搭配词一般都不能与overcome搭配。英语中，“overcome”的常见搭配模式见表18.2。

表 18.2 “overcome”的常见搭配词

	英语中“overcome”的常见搭配词
overcome	problem, obstacle, shyness, difficulty, resistance, shortcoming, weakness, friction, disadvantage, temptation, prejudice, barrier, hardship, jealousy

由于“克服”和“overcome”在英汉语中搭配模式的差异，汉语中的“克服”翻译成英语时就需进行灵活处理，而不应统统译为“overcome”。接下来的问题就是，如何找到“克服”的其他翻译形式。在这方面，汉英对应语料库可提供了重要的参考。我们先按汉到英的方向对“克服”一词在北外通用汉英对应语料库中进行了检索，不同搭配模式的翻译处理如下：

1) 克服现象

(1) 共产党员应和党外一切先进分子协同一致，为着团结全国人民克服各种不良**现象**而努力。Communists should work in harmony with all progressives

outside the Party and endeavour to unite the entire people to **do away with whatever is undesirable.**

(2) 但是我们有信心，我们的党、我们的国家有能力逐步克服并最终消除这些消极**现象**。But we are confident that our Party and our country are capable of gradually **reducing** these negative **phenomena** and eventually eliminating them.

(3) 因此，共产党员不能不自觉地担负起团结全国人民克服各种不良**现象**的重大的责任。Therefore, Communists must consciously shoulder the great responsibility of uniting the entire nation so as to **put an end to** all such undesirable **phenomena.**

在英语中，"overcome"一般不能和"phenomena"搭配，所以，三个"现象"的"克服"译为了：do away with，reduce，put an end to。

2) 克服矛盾

(1) 克服了矛盾，暂时归于一律了；但不久又会产生新**矛盾**，又不一律，又须要克服。The resolution of contradictions results in temporary uniformity, but new **contradictions** soon emerge, which means non-uniformity, and they, in turn, have to be **resolved.**

(2) 而因克服了鱼大水小的**矛盾**，使我们的战争的机构适合战争的情况，我们就将显得越发有力量，我们就不会被敌人战胜，而要最后地战胜敌人。By **resolving** the **contradiction**, which is that of "a big fish in shallow water, and by adapting our war apparatus to the war situation, we shall prove even stronger, and so far from being defeated by the enemy we shall finally defeat him.

在英语中，"overcome"一般不能和"contradiction"搭配，所以，"矛盾"的"克服"译为了：resolve。

3)克服阻力

抓住典型，处理了，效果也大，表明我们下决心克服一切**阻力**抓法制建设和精神文明建设。Dealing with these cases will have the most effect; it will show our determination to **surmount** all **obstacles** in strengthening the legal system and promoting ethical progress.

虽然“overcome”能和“obstacle”搭配，但此处还是译为了“surmount”。

4）克服观念

首先，是由于历史上“好男不当兵”的传统**观念**我们还未能给以有力的克服；First of all, we have not **exploded** the traditional **myth** that “a good man does not join the army”.

我们看到，在此句中，“观念”转换为了“myth”。在英语中，“overcome”一般不能和“myth”搭配，所以，“观念”的“克服”译为了：explode。

5）克服官僚主义

（1）在我们这样大的国家里，有少数人闹事，并不值得大惊小怪，倒是足以帮助我们克服官僚主义。In a large country like ours, there is nothing to get alarmed about if small numbers of people create disturbances; on the contrary, such disturbances will help us **get rid of bureaucracy.**

（2）提高效率，克服**官僚主义**，包括精简机构，还包括其他许多内容。Increasing efficiency and **eliminating bureaucratism** include, among other things, streamlining Party and government organs.

在英语中，“overcome”一般不能和“bureaucracy”搭配，所以，“官僚主义”的“克服”译为了：get rid of，eliminate。

6）克服危险

（1）它能够使我们党得到群众的监督，克服党员堕落腐化的**危险**，及时发现投机分子以及破坏分子而清洗出党；It can keep the Party under the supervision of the masses, so that Party members can **avoid** the **danger** of becoming degenerate and so that careerists and saboteurs can be uncovered and expelled from the Party before it is too late。

（2）所以，斗争是克服投降**危险**、争取时局好转、巩固国共合作的最主要的方法。Therefore, struggle is by far the most important means of **averting** the **danger** of capitulation, of achieving a turn for the better in the situation and of strengthening Kuomintang-Communist co-operation.

在英语中，“overcome”一般不能和“danger”搭配，所以，“危险”的“克服”译为

了：avoid，avert。

7）克服不平衡

太岳、太行区一般应在今年内克服**不平衡**，完成发动群众与组织群众的任务。In the Taiyue and Taihang areas we should **solve** the **problem of uneven development** and arouse the masses and get them organized, for the most part, during this year.

在此句中，"不平衡"并不是一个名词。所以英译时前面加了一个"problem"。在英语中，"overcome"虽能和"problem"搭配，但此处"不平衡"的"克服"还是译为了更常用的搭配词：solve。

8）克服风气

上述两个运动的后期，事前不请示，事后不报告，不按章程办事的无政府无纪律的**风气**，又在抬头，（请中南、华东各省严密注意这些教训，坚决执行请示报告制度，绝对不容许此种坏作风抬头。毛注。）我们已告各地注意克服。In the later stages of the two movements, anarchy and indiscipline began to gain ground when people concerned did not ask higher authorities for instructions or submitting reports afterwards and violated rules and regulations. (Mao's comment: Comrades working in the provinces in south-central and east China are expected to pay close attention to this lesson and see to it that the system of asking for instructions from above and submitting reports afterwards is applied rigorously and that no practices of this undesirable type shall be tolerated.) We have already notified all localities to **check** this **tendency**.

此句中的"风气"，英译时转换为了"tendency"。在英语中，"overcome"一般不能和"tendency"搭配，所以，"风气"的"克服"译为了：check。

9）克服逆转

在这种形势下，争取时局好转，克服时局**逆转**的可能性，还是存在的，中央二月一日的决定是完全正确的。In these circumstances the possibility still exists of **preventing** the situation from **deteriorating** and of changing it for the better; the Central Committee's resolutions of February 1 are entirely correct.

在英语中,"overcome"一般不能和"deteriorating"搭配,所以,"逆转"的"克服"译为了:prevent。

10) 克服思想

(1) 党内要分清理论是非、路线是非,要开展批评和自我批评,互相帮助,互相监督,克服各种错误**思想**。Inside the Party we should distinguish right from wrong in terms of theory and of the Party line, conduct criticism and self-criticism and help and supervise each other in **correcting** wrong **ideas.**

(2) 经过三查三整,克服了党内的混乱**思想**,团结了全党,进一步联系了群众,产生了很好的效果。Through the three check-ups and three improvements, the confused **thinking** within the Party was **clarified**, the Party became united and its ties with the masses were strengthened; in other words, we achieved successful results.

在英语中,"overcome"一般不能和"ideas","thinking"搭配,所以,"思想"的"克服"译为了:correct,clarify。

11) 克服旧毒、污浊、波折、回头路

旧毒,**污浊**,革命进程中的某些**波折**,以及可能的**回头路**,只有斗争和努力才能够克服,而且需要长期的斗争和努力。It is only through struggle and hard work, and over a long period too, that we can **eliminate** the old **evils**, the old **filth**, and **prevent setbacks** or even **reversals** in the revolution.

在英语中,"overcome"一般不能和"evils","filth","setback","reversal"搭配,所以,"旧毒,污浊,波折,回头路"的"克服"译为了:eliminate,prevent。

12) 克服意见

如果对时局的估计和任务的提出发生过左过右的**意见**,而不加以说明和克服,那也是绝大的危险。It would be perilous not to point out and **correct** any "Left" or Right **deviations** in appraising the situation and in defining our tasks.

在此句中,"意见"英译时转换为了"deviations"。在英语中,"overcome"一般不能和"deviations"搭配,所以,"意见"的"克服"译为了:correct。

13）克服贪污、行贿、盗窃、东西

对**贪污**、**行贿**、**盗窃**以及其他乌七八糟的**东西**，人民是非常反感的，我们依靠人民的力量，一定能够逐步加以克服。The people detest **graft**, **bribery**, **theft** and other dirty **practices**. I have no doubt that if we rely on the strength of the people, we shall be able to gradually **eliminate** those practices.

在英语中，"overcome"一般不能和"graft"，"bribery"，"theft"，"practices"搭配，所以，"贪污，行贿，盗窃，东西"的"克服"译为了：eliminate。

14）克服缺点、错误

（1）所有这些**缺点**必须加以克服。We must do away with all such **shortcomings**.

（2）要相互容忍，相互谦让，相互帮助，相互补充，包括相互克服**错误**和**缺点**。You should be tolerant and generous with each other, help each other, complement each other's thinking and help **correct** each other's **mistakes** and **shortcomings**.

（3）我方反是，我之优点可因我之努力而加强，**缺点**则因我之努力而克服。On the other hand, our advantages can be enhanced and our **shortcoming** **remedied** by our efforts.

在英语中，"overcome"能和"shortcoming"搭配，但以上3句中的"克服"还是采用了多种译法：do away with，correct，remedy。但"overcome"不能和"mistake"搭配，所以第2句中"克服"的翻译采用了"correct"。

将汉英对应语料库中汉到英方向的"克服"的英语对应词小结见表18.3。

表18.3 "克服"的英语对应词（汉到英方向）

	英语对应词
克服	do away with, surmount, reduce, explode, resolve, surmount, get rid of, avert, solve, check, prevent, eliminate, correct, avoid, remedy, put an end to, clarify
	共计：17个

同时，还可在汉英对应语料库中按英到汉的方向对"克服"进行检索，检索行如下：

（1）Fortunately Scarlett had come into contact with the two men very

little, for Will had persuaded her to let him handle the trading while she managed the plantation. In his mild-tempered way, Will had **straightened out** several difficulties of this kind and said nothing to her about them. 幸好思嘉很少同这两个人发生什么联系，因为威尔早已说服她让他来管理买卖上的事，而她自己只管理农场。威尔用他那种温和的办法克服了好几种这一类的困难。并对她什么也没有说。

(2) Greater things than the Doctor had at that time to **contend with**, would have yielded before his persevering purpose. 即使当时医生要克服的困难比现在还要大得多，在他那坚持不懈的努力之下困难也是会退让的。

(3) "If I do, she'll **outgrow** it and forget about it." 只要我这样做，她就会逐渐克服它，把它忘了。

(4) Knowing your natural temper better than I, he could the better judge what arguments to use, whether of tenderness or terror, such as might **prevail over** your hardness and obstinacy, insomuch that you should no longer hide the name of him who tempted you to this grievous fall. 由于他比我更了解你的秉性，他应该是个更合格的法官，他更清楚应该选用什么样的刚柔相济的辞令，来克服你的桀骜不驯；以使你不再隐瞒那个诱惑你如此堕落的人的姓名。

(5) Mapping the position of genes on chromosomes gives today's biotechnology researchers a huge advantage over their predecessors in discovering the molecular activities involved in diseases and devising new, more effective therapeutic interventions to **counteract** them. 确定染色体上的基因位置，给今天的生物技术研究者来比他们前辈更大的便利，例如发现疾病的分子活动和设计新的更有效的治疗介入方法克服疾病。

(6) We know now, the natural antipathy you strove against, and **conquered**, for her dear sake. 现在我才明白了你为她的缘故跟发自天性的憎恶作了多少年斗争，并且克服了它。

(7) But, few cared enough for that to think of it a second time, and, in this matter, as in all others, the common wretches were left to **get out of** their difficulties as they could. 可是注意到这类事件并加以考虑的人却很少。因而在这件事上也跟在别的事上一样，普通的穷苦百姓便只有自行努力去克服困难了。

(8) He, on this part, had pretty well **fought** his weakness **down**. 至于希尔弗，他已经比较成功地克服了自己一时的怯懦。

将汉英对应语料库中英到汉方向的“克服”的英语对应词小结见表 18.4。

表 18.4 “克服”的英语对应词(英到汉方向)

	英语对应词
克服	straighten out, contend with, outgrow, prevail over, conquer, counteract, get out of, fight down
	共计:8 个

下面,将汉英对应语料库中英到汉与汉到英两个方向的“克服”的对应词总结见表 18.5。

表 18.5 “克服”的英语对应词(双向)

	英语对应词
克服	do away with, surmount, reduce, explode, resolve, surmount, get rid of, avert, solve, check, prevent, eliminate, correct, avoid, remedy, put an end to, clarify, straighten out, contend with, outgrow, prevail over, conquer, counteract, get out of, fight down
	共计:25 个

这样,通过汉英对应语料库中翻译实例的考察,学生就会加深对汉英两种语言中搭配模式差异的了解,并且找出汉译英中解决搭配模式差异的具体方案。Pym (2003)提出了翻译能力的最简方案,包括两部分:1)一种原文本生成不止一种,而是一系列的可行的目标文本(TT_1, TT_2—TT_n)的能力;2)从上述系列目标文本中,迅速、正确、有信心的仅选择一种可行的目标文本的能力。该方案对翻译中的搭配教学有重要意义。由于汉英搭配模式的差异,译者首先应能够生成一系列的搭配模式或搭配变体,然后从中选择出最适合当前语境的搭配选项,以保证译文中的搭配得当。

此外,由于英语不是我们的母语,所以我们认为汉译英中目的语英语的单语母语语料库对汉译英也有一定帮助作用。在汉译英中,BNC 是较好的单语母语参考语料库。下面是我们对“eyes”一词在 BNC 中进行检索得到的部分主谓、定中和动宾搭配。

1) 主谓搭配

their mouths fell open and their <w NN2> eyes **popped** with astonishment.

but & hellip; & equo; Sladen's <w NN2> eyes **searched** for comfort.

his strong dark face and amber <w NN2> eyes **danced** between her and the

shoulders like a caress. Blue <w NN2> eyes **smiled** from a face that,
ed the ice in his tone and his <w NN2> eyes **surpassed** it. She face
she said automatically, but then her <w NN2> eyes **flicked** to his in alarm
& equo; she said, anger making her <w NN2> eyes **flash.** & bquo; Somewhe
y second skin. & equo; His <w NN2> eyes **crinkled** with amusement.
cusing me? & equo; he asked, his <w NN2> eyes **laughing** at her. No
impossible in any dream. Her <w NN2> eyes **flew open** as Roman kissed
, and he went very still, his <w NN2> eyes **leaping** with fire as he de
h you, & equo; Claudia said, her <w NN2> eyes **darting**, from one single
About time, & equo; he said, his <w NN2> eyes **taking in** the picture she
e under the flimsy nightgown, his <w NN2> eyes **caressing** her as blatantly as
o; He smiled again, his <w NN2> eyes **hooding** as he looked at he
urned like fire. Luke's <w NN2> eyes **devoured** her. & bquo; D
& equo; he said. Her <w NN2> eyes **strayed** involuntarily to the
ch way they would go. His <w NN2> eyes **glazed** on the swaying barley
ansparent, from the way his amber <w NN2> eyes **lingered** on every inch of

2) 定中搭配

of brown dust, and their **yellow** <w NN2> eyes a slightly inebriated, glazed
rkness of hair, with big **golden-brown** <w NN2> eyes that saw everything as wildly
ves for a glance from his **magnificent** <w NN2> eyes. He wouldn't look
th the same light, bright **aqua-marine** <w NN2> eyes that looked out from her
ined fingers and the watery **bad-tempered** <w NN2> eyes. Oh, she was g
aria looked straight at him with **hard** <w NN2> eyes. & bquo; I still hate
true, & equo; she insisted, her <w NN2> eyes still **stormy.** & bquo;
Not cocky young men with roving <w NN2> eyes and a nice taste

in girls

hidden. His **lazy**, **supercilious** <w NN2> eyes, too, managed their affect

features and long, **indifferent lapis** <w NN2> eyes, belonged to a statue rath

white-faced with huge **puffy red-rimmed** <w NN2> eyes, his thin hair plastered t

's really pathetic. His little <w NN2> eyes are completely **raw**, he can

the gown round her roughly, his <w NN2> eyes as **cold** as stones.

3) 动宾搭配

to sting and then **burn at** his <w NN2> eyes so that he was finally forced

joined with him, and **rolled** her <w NN2> eyes, having seen the awful pun

then she **dabbed** her swollen <w NN2> eyes with the tissues Belinda

我们看到,"eyes"一词在 BNC 中的许多搭配模式不是常见的,不是词典中能找到的,也不是我们所能想到的,反映出了英语语言中该词的搭配模式特征。因此,汉译英中,我们需要对目的语英语中的搭配模式进行深入了解。在这方面,BNC 是重要的参考语料库。

18.4 小结

搭配是汉译英教学中的一个大问题。汉译英教学中,搭配模式的转换是否得当往往依赖授课教师的直觉。但一个人的语言直觉往往又是有限的,不能把握过多的搭配形式。本章通过语料库对搭配模式进行调查,发现汉英对应语料库和目的语为英语的单语母语语料库可为汉译英中的搭配教学提供重要的资源和参考。一方面,可为译者提供搭配的多种选择形式,开阔译者的搭配视野;另一方面还可为译者选择最佳搭配模式提供重要参考。因此,语料库是译者提高自身翻译能力的重要途径。

第 4 部分

语料库翻译学及相关应用问题

第 19 章

计算语言学与翻译研究

通过对双语语料的分析，本章从形式和语义两方面探讨计算语言学方法对翻译研究的作用和启示。在语言形式方面，通过文本特征统计信息研究了双语文本的基本形式特征，通过赋码语料研究了双语文本的深层句法信息。在语言意义方面，运用潜语义分析方法计算语句意义间的相似度，语句计算的结果与人工评判的结果相差无几，表明计算语言学的意义分析方法十分有效，对判断多个译文间、译文与原文间的“形似”与“神似”有重要意义。

19.1 引言

计算语言学(computational linguistics)是从计算的角度对语言进行研究的一门学科。俞士汶等(2003:2)曾给计算语言学下了一个相对严谨而概括的定义：“计算语言学指的是这样一门学科，它通过建立形式化的数学模型来分析、处理自然语言，并在计算机上用程序来实现分析和处理的过程，从而达到以机器来模拟人的全部或者部分语言能力的目的。”计算语言学与翻译有着不解的渊源，实际上就是起源于 20 世纪 50 年代的机器翻译。虽然当时的语言自动处理咨询委员会(ALPAC)的报告对机器翻译采取了否定的态度，指出机器翻译研究遇到了难以克服的“语义障碍”(semantic barrier)。然而，计算语言学作为一门交叉学科并未因此而停滞不前。相反，随着人工智能学科的蓬勃发展，计算语言学运用自然语言处理方法对翻译研究起到了越来越重要的作用。其实，翻译与计算语言学两者间本来就联系紧密，它们的处理对象同为自然语言，所不同的是，传统翻译中语言处理的主体是人，而计算语言学中语言处理的主体则是人编写的计算机程序。计算语言学中运用自然语言处理方法所获得的信息可以借鉴到翻译研究之中，而惯常的翻译研究方法也可以用来指导计算语言学的自然语言处理实践，两者有很多共融共通的方面。

19.2 计算语言学在翻译中的应用

翻译是两种语言之间的转换活动，前苏联翻译理论家费道罗夫认为“翻译就是用一种语言把另一种语言在内容和形式不可分割的统一中业已表达出来的东西，准确而完全地表达出来”(林煌天，2005：181)。费道罗夫在他的翻译定义中强调了语言形式和内容之间的关系。王克非(1997：12)：指出：“翻译中对原作要力求忠实于内容，兼顾其形式，顺畅地表达。”贝尔(1991：5)也认为翻译实际上就是把源语中的东西用目的语表达出来，并尽量保持其语义和文体的等值。在上述关于翻译的论述中，内容和形式被认为是翻译定义的两个重要组成部分，任何语言都包括形式和意义两个部分，因此本章拟从语言形式和语言意义两个方面论述计算语言学在翻译研究中的作用。

19.2.1 语言形式方面

19.2.1.1 文本语言特征统计信息

计算语言学采用自然语言处理的方法提取文本特征统计信息，例如形符(token)和类符(type)及其比率、标准化的形符类符比(standardized type-token ratio)、频率、平均词长、句长、段落数目等。下面以鲁迅先生的作品《风筝》为例试做分析(译文摘自刘士聪(2002)编写的《汉英英汉美文翻译与鉴赏》)。

表 19.1 和表 19.2 的数据均由语料检索软件 Wordsmith 4.0 对《风筝》的中英文分别进行词频统计所得。考虑到中英文的差异，我们在对汉语进行处理之前，用中科院汉语词法分析系统(ICTCLAS)对汉语进行了分词处理，增强了中英文的可比性。表 19.1 的数据显示，汉语原文和英文译文的文本特征统计信息存在较大差异，英语的形符数和类符数都明显高于汉语。笔者曾以经过 ICTCLAS 分词后的《红楼梦》及其两个英文译本为例进行查询，发现《红楼梦》的原文词语数目为 657 805 词，而霍克斯译本的词数达到 1 140 020，超出汉语原文 73%，而杨宪益译本的词数也达到 857 338，超出原文 30%。王克非(2003)通过基于汉英对应语料库的考察也发现译文有扩增现象。另外值得注意的一点是，两种语言在句子数目上也存在差异。这一结果从一个方面表明了英汉两种语言的句法特征。英语句式一般层次感和规律性较强，句子写得过长常常会使得句法层次感削弱，显得晦涩难懂，因而相对较少出现，而汉语句子中各成分之间的关系则常常较为松散和自由，并列句往往更为常见，这使得汉语句子可以容纳的信息比英语更多。因此，在汉译英的过程中，原文一个句子在目标语中常常需要分解成两个甚至更多的句子。相反，在英译汉的过程中，原文两个或者更多的句子常常可能合并成目标语中的一个

句子。

表 19.1　原文和译文文本特征统计信息

特征统计项/语种	汉语原文	英文翻译
形符	832	1 056
类符	393	439
类符形符比	47.236	41.572
句子数目	41	51
平均句长	20.29	20.71

表 19.2　汉语原文和英语译文中居前的十个词及其频率

汉语词语	的	我	了	他	是	着	在	风筝	很	一
频率	72	37	24	24	18	15	13	12	10	9
英语词语	the	and	be	of	I	a	to	in	with	my
频率	59	41	41	40	34	33	29	20	20	17

表 19.2 列出汉语原文和英文译文中出现频率最高的前十个词语。比较原文和译文中的高频词语可以发现，无论在英语还是汉语中，虚词的频率远远高于实词，只是虚词的种类有所不同，而这种不同恰恰从某种程度上体现了英汉两种语言之间的差异，了解这些差异对翻译研究具有重要意义。根据高频词的使用情况，我们至少可以看到英汉语间以下几方面的差异：1)所属关系的表达方式。由于汉语词语几乎没有形态变化，常常通过助词来表达代词或名词的属格关系，因此汉语中助词"的"出现的频率很高(如"我的"、"风筝的"等)。英语则不同，代词的属格常通过形态变化(如 my 可以看作为 I 的属格)实现，名词的属格则常常通过在两相关名词间加介词 *of*(如 roof of the house)来实现；2)时态的表现方式。汉语依靠"了"、"着"、"过"等助词表现时态，而英语则通过动词的形态变化表现时态；3)名词限定词(determiner)的差异。汉语中名词之前的限定词远远不如英语中(如 the，a，my 等)那么常用；4)介词的差异。英语的介词不仅使用频率高，数量也比汉语的多。汉语介词"在"(如"在……上"、"在……下"、"在……前"、"在……后"等)适用于多种场合，而英语中众多介词之间常常有较为明显的分工；5)指代(reference)问题，汉语倾向于省略指称词，而英语则出于语法和语义的需要不能省略，例如汉语原文中用于指称事物的"它"只出现了一次，而英语译文中的"it"却高达 14 次。英汉语间以上各方面的差异对翻译研究和翻译过程中目标语的选择等都具有重要意义。

19.2.1.2 深层次文本特征挖掘

对翻译文本的处理,研究者不能只停留于表面的文本特征统计信息,还需要借助语言处理工具对现有文本做进一步的研究。在语言形式方面,除了上述提及的文本信息,研究者还可以对原文和译文的词法、句法进行研究。

对文本的词法和句法进行研究常常需要借助计算语言学领域的词性标注软件(POS Tagger)进行。常用的英语词性赋码软件有 Brill Tagger,CLAWS(Constituent Likelihood Automatic Word-tagging System)等。CLAWS 是兰开斯特大学从 20 世纪 80 年代开始研制的一种基于概率的词性赋码工具,CLAWS 宣称该软件赋码的准确率达到了 97%,英国国家语料库(British National Corpus)就是运用 CLAWS 进行赋码的。汉语的词性赋码工具主要有上面介绍过的中科院汉语词法分析系统(ICTCLAS)以及东北大学自然语言处理实验室研制的 NEUCSP 分词词性标注工具等。

依旧以鲁迅先生的《风筝》的翻译为例。我们用 NEUCSP 对汉语原文进行分词标注,英语译文则使用 CLAWS 进行词性标注。进行词性赋码后分别把原文和译文中的汉语词语和英文词语使用自行编写的程序删去,只保留词性赋码。这些词性赋码实际上就是原文文本和译文文本的形式句法特征库。以原文和译文其中一句为例:

(1)a 原文:我忽然想起,似乎多日不很看见他了,但记得曾见他在后园拾枯竹。

(1)b 词性标注后的原文:我/r 忽然/d 想起/v,/wp 似乎/d 多日/nt 不/d 很/d 看见/v 他/r 了/u,/wp 但/c 记得/v 曾/d 见/v 他/r 在/p 后园/n 拾/v 枯/a 竹/n。/wp

(2)a 译文:It occurred to me that I did not seem to have seen him for days, but remembered having once met him in the back yard collecting dry bamboo sticks and twigs.

(2)b 词性标注后的译文:It_PPH1 occurred_VVD to_II me_PPIO1 that_CST I_PPIS1 did_VDD not_XX seem_VVI to_TO have_VHI seen_VVN him_PPHO1 for_IF days_NNT2,_, but_CCB remembered_VVD having_VHG once_RR met_VVN him_PPHO1 in_II the_AT back_NN1 yard_NN1 collecting_VVG dry_JJ bamboo_NN1 sticks_NN2 and_CC twigs_NN2 ._.

以上词性标注使得句法分析成为可能。我们运用自行编写的词频统计软件结

合 Wordsmith 4.0 对词性赋码进行检索，结果见表 19.3：

表 19.3 原文和译文中居前的十个词类范畴

原文	V（动词）	U（助词）	N（名词）	D（副词）	R（代词）	A（形容词）	C（连词）	P（介词）	NT（时间名词）	M（数词）
频率	191	129	114	99	87	53	25	22	20	14
译文	N（名词）	I（介词）	V（动词）	P（代词）	AT（冠词）	C（连词）	J（形容词）	R（副词）	APPGE（所有格代词）	D（限定词）
频率	186	141	136	101	94	67	62	38	26	21

原文中，动词的数量最多，而译文中名词的数量最多，表明原文中一部分动词译成了名词，例如：

(3) 灰黑色的秃树枝<u>丫叉</u>于晴朗的天空中。
An <u>**entanglement**</u> of dark-gray bare branches against a sunny sky.

还有一部分动词翻译成了形容词，例如：

(4) 他只得<u>**张着**</u>小嘴，呆看着空中出神。
His tiny mouth **open** in a gape, watching upward.

其次，汉语原文有 25 个连词，而英语译文有 67 个连词。在 CLAWS 标注集中，连词共分为 7 小类，在本章中出现了 5 小类，类别数量如下：CC(并列连词)45 个、CST(连词 that)9 个、CS(从属连词)6 个、CCB(转折并列连词)5 个、CSA(连词 as)2 个。从连词数量的对比可以看出译者在汉译英中倾向于用显性化的连接手段来表达原文中隐性的成分。显性翻译被认为是翻译中的一个普遍倾向(Chesterman 1997:71)。柯飞(2005:307)也指出"汉译英时会比英译汉时更多地呈现形式上的'显化'，而英译汉时则呈现更多的模仿式的'显化'"。除了对一元词语(unigram)进行分析研究外，研究者还可以对二元词语(bigram)和三元词语(trigram)直至整个篇章的句法结构进行研究分析。在本例中，汉语原文中出现频率最多的二元词语和三元词语分别是：D(副词)＋V(动词)和 D(副词)＋D(副词)＋V(动词)，分别为 41 次和 12 次；而在英语译文中出现频率最多的二元词语和三元词语分别是：AT(冠词)＋NN1(单数名词)和 AT(冠词)＋NN1(单数名词)＋IO

(介词 of),分别为 50 次和 18 次,基本与表 19.3 中的词性赋码频率相吻合。当然除了对词元进行研究之外,也可以对句法结构(如被动语态、定语从句等)进行研究。

19.2.2 语义意义方面

计算语言学在语言意义方面研究的重要应用之一是对机器译文和学生译作的评测。对于机器译文的自动评价目前主要集中于两种评测方法:一个是 IBM 提出的 BLEU(Bilingual Evaluation Understudy)的评测方法,另一个是 NIST(National Institute of Standards and Technology)评测标准。它们都是从文本的 N 元词语(N-gram)出发,比对测试文本和参照文本之间的 N 元词语的相似程度,计算出被测文本和参考译文之间的距离。BLEU 默认匹配四元(4—gram)以下的相似度,因为超出四元的连续词语对自动评价的意义不大。计算 BLEU 数值的公式(Papineni & Roukous,2002:311-318)为:

$$Bleu = BP \times exp \sum_{n=1}^{N} w_n \log P_n$$

其中 BP(Brevity Penalty)为长度罚分;w_n 为各阶匹配正确率的权重系数。P_n 为 N 元匹配时的分数。

NIST 是在 BLEU 标准基础上提出的一个评测方法,它考虑了信息权重的因素。这两种评测方法实质上是对被测译文和参考译文在 N-gram 层次上进行匹配,它们得出的相似度仅仅是完全匹配了的译文的重合度(overlap),这样的方法存在着一个致命的弱点,即在自然语言的使用中,存在着很多近义词项,无法通过这种方法来计算语句之间的相似度。例如,任何自然语言中都存在着大量的多词同义(synonymy)和一词多义(polysemy)现象,且语言单位间的相似程度并非非此即彼,以 0 和 1 两个极端值来代表语言单位间的相似度显然有悖常理。在当今计算语言学领域,自然语言中语言单位间的相似度常常用 −1 与 1 之间的若干数字来表示,关系近的相似度应该接近于 1,而远的则趋向于 0,直至 −1。语义相似度量化测量的方法之一是借助于信息检索(information retrival)领域的"潜语义分析"(latent semantic analysis,LSA)。

潜语义分析运用数据统计以及数学分析的方法,通过观察词项在同一文本中出现的相对频率来计算语义的相似度。潜语义分析超越了只停留在语句表层的语义分析方法,它把自然语言中的每个文档视为以词语为维度的空间中的一个点。同样,每个词语被视为以文档为维度的空间中的一个点,通过点与点的交集构成一个矩阵(matrix),而这个矩阵表征的是整个一批文档之间的深层,即潜在的语义空间。Manning 等(2005:344)曾有过一段通俗的论述:"潜语义分析就是把同现词条映射到同一维空间上,而非同现词条映射到不同空间上。从这个意义上来讲,即使

查询字段和文档没有共同词条，它们的余弦相似度可能仍然很高，其中的缘故就是同现分析。同现词条被认为是语义相关的。"通过计算词语的权重(weighting)和奇异值分解(singular value decomposition)对矩阵进行降维，去除不应有的噪声，保留语义空间中的主要框架。在国外，Laham(1997)、Landauer，Foltz 和 Laham (1998)、Turney(2005)等用 LSA 的方法做过不少与语义相关的研究，他们得出的结论显示，通过 LSA 分析句子间的相似度，可以有效地避免自然语言中多词同义和一词多义问题，提高分析的准确度。经反复验证，以 LSA 计算的句子间的语义相似度与人工对意义判断的结果极为接近，充分说明了 LSA 对语句的相似性有着很强的预测作用，使用潜语义分析法分析语句间的相似度有着传统分析方法无可比拟的优势。

运用潜语义分析可以分析译文的语义内容并且能够把译文依据意义相似程度排序。下面的一句译文就取自一个由 150 句子构成的小型语料库，语料库中共包含 30 个句组，每组由相似度各异的五个句子组成。

(5) 原文：国家利益就是使得国家资源物尽其用。

1. The national interest lies in all the nation's resources being put to the most advantageous use possible.

2. The excess land was mostly sold to existing tenants.

3. Using all the nation's resources as advantageous as possible represents the national interest.

4. Our national interest was best served by removing the Saddam regime, given that we were prepared to dedicate the time and resources to post-war Iraq.

5. The national interest consists in putting all the nation's resources to the most advantageous use.

我们通过运用潜语义分析对建成后的矩阵进行奇异值分解(SVD)，然后计算列向量之间的夹角余弦值。这个值介于－1 和＋1 之间，＋1 表示文本完全重合，－1 表示语义内容完全不相关。

表 19.4 SVD 语句相似度计算结果

	第一句	第二句	第三句	第四句	第五句
向量夹角	1	1.843087	0.288002	1.438137	0.036659
夹角余弦值	1	－0.26894	0.958813	0.132271	0.999328

在表19.4中，以第一句译文为基准，与第一句意义最相似的是第五句，其次为第三句、第四句，最不相似的是第二句。其中，很明显第一句、第五句和第三句意义非常相近，属于同一个意群，而第四句则与前三句在语义上有着较大的距离，第二句则与前面的句子在语义上毫无关联，其数值为负数。语句计算的结果与人工评判的结果相差无几，表明计算语言学的意义分析方法十分有效，对判断多个译文间、译文与原文间的"形似"与"神似"有重要意义。

19.3 小结

计算语言学可以通过对双语语料库的有效分析，从形式和语义两个主要方面为翻译研究、译文质量评价和机器翻译实践等提供大量的语言信息。该学科中复杂而高效的逻辑运算、数学分析以及高度形式化的计算模型，可以为翻译研究的深入提供重要的量化资源。黄立波、王克非(2006:37)指出翻译研究不能仅仅局限于语内对比，并且提出"基于语料库的(翻译)普遍性研究需要打破类比语料库研究模式的垄断，重新将源语文本纳入其研究的视野"。借助计算语言学的知识和工具，我们可以实现这一目标，把源语文本纳入翻译对比研究的视野之中。借鉴计算语言学的方法和理论，翻译研究能够拥有更为开阔的视野，开拓更为广阔的空间。

第 20 章

汉英对应语料句对检索

汉英对应语料在经过对齐标注校对入库之后，便转化为可管理的数据库系统。为使这一宝贵的语言资源能够更好地服务于翻译教学与翻译教学，我们制作了一个前端用户访问后台数据库的服务接口，通过较好的人机交互界面，提供一个对双语语言事实进行更多复合条件查询的检索系统。本章简要介绍这一系统的设计开发思想及其应用。

20.1 引言

当前从事机器翻译的研究机构为建立较好的语言模型和翻译模型，分别在搜集整理双语语料，建立起越来越大的双语对应语料库。有的更进一步利用网络爬虫等技术手段从互联网上获取更多翻译语料来丰富语料库。然而建库不是最终目的。为了充分利用这些语言资源，于是更多检索工具也开发问世了。由于版权限制等各方面原因，目前能够下载或在互联网上运行的双语语料库检索系统并不多见。即使部分能够运行，检索界面也往往比较简单，比如只提供简单的中英文字词的语言内容查找，有些甚至没有进行基本的英语词形还原，因而使得由于检索词和语料中的词语不一致造成的漏检现象频繁发生。同时对检索的元属性限制条件，譬如翻译方向的汉译英、还是英译汉；语体特征的书面语还是口语等属性检索基本没有提供。由于设计上的这些缺憾，使得以机器翻译为主的语料库检索对面向翻译研究者和学习者的贡献打了一个折扣。

近年来，在国家社会科学基金和教育部人文社科重点研究基地重大项目的支持下，我们已经建成了当今世界上经过专业人员校对、规模最大的汉英对应语料库（王克非 2004a）。本着切实为翻译教学与研究服务的目的，我们设计了一个用户检索系统，希望通过简单的人机交互接口，使得更多用户能够方便地获取利用这一宝贵的语言资源。最近我们在北京外国语大学中国外语教育研究中心的“中国外语教研网”发布了一个 Web 版的原型检索系统，以听取更多用户的意见和建议。

20.2 语料库的数据库转化设计

语料库作为海量语言事实材料的集合，一般来说，数量相对较大。采用一般的文件处理系统难以管理。譬如我们创建的汉英对应语料库的总字数就以千万计。搜集本身不是目的，更关键的是加工利用。对语料的处理既可以自行设计数据结构或采用专门软件进行管理。

为简化开发时间，我们在汉英对应语料库平台中，采用关系型数据保存、管理、检索语料。因为数据库本身具有良好的数据存储及管理与维护功能，方便后续语料库不断追加；数据库还提供较好的检索二次开发接口设计，能在此基础上方便快捷地定制特定的查询处理。在本项目中，双语对应语料库实际上设计为数据库中的一个表(table)。该表是由一个个汉英对应的句子及其他与句对相关的元属性共同构成的记录(record)组成。每条记录都包含有如表 20.1 所示字段(field，属性)的信息内容。

表 20.1　数据库字段设计及说明

字段名	类　型	备　注
英文句	字符型(5 000)	一个英文句子不超过 5 000 字母
汉语句	字符型(2 000)	一个汉语句子不超过 2 000 汉字
源语言	字符型(1)	当前句子的源语言(英语或汉语)
作者	字符型(50)	当前句子的作者名
篇名	字符型(50)	当前句子所在文本的篇名
类型	字符型(1)	当前句子所属的类型(取规定值之一)
文体	字符型(1)	当前句子所属的文体(取规定值之一)
语体	字符型(1)	当前句子所属的语体(取规定值之一)
时代	字符型(1)	当前句子所属的时代(取规定值之一)
段落 ID	数值型	当前句子所在段落相对全文偏移量
句子 ID	数值型	当前句子相对段落的偏移量
文件名	字符型(50)	当前句子所在文件名

在如表 20.1 所示的数据库字段的规划中，最主要的是“英文句”和“汉语句”，分别存入互有对应关系的完整英语和汉语句子。英汉句子还分别进行了词类赋码，数据库记录中实际存放的是带有赋码标记的字符串。其他字段记录的是当前

汉英句子对所属语篇的文本属性(源语言、文体)或依附属性(如作者名、篇名),譬如从该句源语言的属性可以推知是“英译汉”还是“汉译英”等不同的翻译过程;从文体属性可推知是文学体还是非文学体等。对语料库检索或过滤时,可以根据这些字段属性给予各类综合条件的限制,从而得到所需结果。

20.3 用户查询检索接口设计

将原始双语语料的对齐是创建双语对应语料库的基础,标注是对其加工的深入,校对能够保证其质量,入库使得语料库成为可管理的数据库系统。检索系统是终端用户与数据库之间的接口。通过这一界面,用户能够比较简捷、方便、有效地获取对语料库不同需求的访问。

20.3.1 检索问题

数据库系统管理维护等操作可以通过结构化查询语言(Structured Query Language,SQL)完成。虽然 SQL 对数据库管理员等专业人士来说并不困难;但我们不能对从事翻译教学与翻译研究的用户 IT 技能要求太高。为此系统提供了一个图形化的用户查询界面。通过这个界面,用户直接可以键入待查询的内容,同时选择不同元属性的限制条件,系统在后台接受用户输入后自动生成合适的 SQL 命令语句,并对数据库执行该命令返回用户希望的结果。该命令既可能是对语言翻译本身的简单查询,也可能是结合更多元属性限制条件的复合查询。

所谓检索是从数据库中把满足用户指定限制条件的某些特定记录找出来。从上文的数据库结构设计中可以看出,所有表示元属性的字段取值都是从固定集合中取值,并且取值是唯一的,譬如对任意一个汉英翻译句对所处文本的写作年代必定是“20 世纪前”、“20 世纪上半叶”和“20 世纪下半叶”中的一个,在数据库中“时代”字段取值编码必定为{1,2,3}之一。因此只要熟悉编码规则,或参考帮助文件,就可以简单直接地使用这些元属性取值作为限定条件。针对翻译现象本身的检索式则难度要复杂得多。譬如查询语言不一、可能是汉语或是英文,抑或是两者混合;查询长度不一,既可能只有一个词语,也可能是若干个短语;查询主体不同,可能是词,也可能是码,或是词加码的组合;查询性质不同,可能是需要在目标句对中出现的词语,也有可能是不需要在目标句对中出现的过滤词。

词类赋码是一个封闭的标记集,这些标记可以悉数枚举,检索起来并不构成困难。用户输入的查询词与实际在语料库中出现的词可能由于词形的不一致而造成漏检或误检。有鉴于此,我们在本系统中分别采用模糊检索和词形还原来解决此类问题。

20.3.1.1 解决汉语分词不一致的模糊检索

为满足用户对词类的检索，数据库中存储的汉语句子都是已经过分词和词类赋码的文本，汉语词之间都已经添加了空格，并且每个词之后都缀上了相应的词类码。因此有可能出现语料库中分词系统切分出来的词语与普通用户输入词语不一致的情况。譬如假定用户输入字符串“中国人民解放军”，认定这是一个词，因此没有以空格隔开。而在语料库中，串“中国人民解放军”实际被切分成“中国_np 人民_n 解放军_n”等三个“词＋码”的子串拼接形式。这样将检索不到用户希望得到的句对。由用户和分词系统对词的认识不一致，采用字符严格匹配的方式容易导致漏检。

我们的处理策略是，首先将查询字符串拆解，汉语以字，英语以单词为基本单位；对各基本单位后采用正则表达式加接 0 或 1 个词类编码形式，构造新的基本单位；然后再拼接各基本单位构造新的查询表达式，这样能够检索得到与原查询串基本单位序列相同并且字形一致但可能被打散或整合的汉语句子，以防止对分词理解不同造成的漏检。应该注意的是，这种处理也有可能误检出基本单位字形相同并且排列相同但实际并非用户希望检出的句子。如假定要检索“华人”，切分查询串为字后，语料库中的“中华人民共和国”将被检出。因此模糊检索是强调检出率优先的可靠选择，适用于用户追求全面结果，不希望有任何遗漏，然后再人工逐一检查这样一种后处理。

20.3.1.2 解决英语词语匹配不一致的词形还原

作为一种形态相对比较丰富的语言，英语的词形变化可能导致字符串匹配的困难。譬如可数名词有单复数的屈折变化；动词也有现在分词、过去分词及过去完成式等各种不规则变化形式。因此，对用户输入的英语查询串，需要有个词形还原形式。假定输入的是词典词的原型形式，则需将其在实际文本中可能呈现的其他屈折变化形式都能检索出来。譬如待查询的是原型动词“go”，则需要将实际例句中可能出现的“went/gone”等其他变体形式也包罗进来。

我们建造了一个包含 4 万个词语及其相应不规则形式的词表，包括词典词的原型及其衍生出来的名词单复数、动词时态还有形容词比较级及最高级等形态变化。采用查词表的方式对输入查询串中的英语词尝试进行扩展。其中新扩展的词语与原查询串中的原型词之间保留逻辑上的“或”关系，即在语料库的句子中只要能找到原词或其扩展词的任意其一，都将视作对查询条件的满足。

20.3.2 Web 和桌面应用程序设计

目前我们实现了两种检索方式的雏形，一是网络查询，是桌面查询。前者通过 Web 浏览器联网注册成为用户后，登录指定网站，根据界面指导完成查询工作；后者采用光盘将系统安装在用户计算机桌面使用，通过用户图形界面检索。

网络版检索系统在后台选用开源数据库 MySQL 作为语料库的管理查询工具，前端采用脚本语言 Php 制作用户交互式的 Web 查询界面。通过文本框接受用户输入，下拉框的设计用来接受用户对特定属性的限制，构造一个可直接对数据库检索的查询语句，并将执行结果输出到用户浏览器页面。其中用户指定的查询词语将以高亮形式凸显。这是典型的瘦客户端工作方式。用户自身的计算机无需安装特别的程序或其他系统，只要浏览器能联网就能访问建好的双语对应语料库系统。相应的缺陷是对服务器负担较重，如果某一时段的并发访问过多则有可能造成服务器宕机；另外对用户的网络连接速度要求也较高，否则有可能因访问超时浏览器停止工作。

至于单机版检索系统，我们曾在项目组内部采用 C# 和嵌入式数据库 Sqlite 在微软 DotNet Framework 平台下完成了一个集语料校对、数据入库等内部处理和查询检索等外部工作为一体的开发平台。查询界面与 Web 方式基本相同，也是通过文本框和下拉框实现用户输入查询串和选择元属性，经由程序将用户输入内容翻译为等价 SQL 命令语句，将结果高亮呈现在浏览器 webBrowser 控件中。

图 20.1 和图 20.2 分别显示 Web 和桌面应用下的查询实例。其中，Web 方式显示的是用户查询“克服 困难 difficulty overcome”，系统设计查询串的所有原子之间缺省为逻辑“并”的关系，因此所有汉语句子带有“克服 困难”并且对应的英文句子带有“overcome difficulty”等词形的实例将作为结果输出，同时这些查询词语将高亮显示，其中英语词可能的其他变体形式，如 overcame 和 difficulties 等视作查询原词。桌面应用程序显示的是用户查询“克服 困难-overcome difficulty”，根据系统设计前加“-”的查询词语被视作排除词，将被过滤，因此所有汉语句子含有“克服 困难”而相应的英文句子带有“difficulty”，没有“overcome”的句对将被检索出来。采用这种限制方式，能够检索出汉语“克服困难”在“困难”翻译成“difficulty”的时候，“克服”还可以用除 overcome 之外英语翻译。



图 20.1　Web 方式用户查询



图 20.2 桌面版用户查询

20.4 汉英翻译单位的对应检索

在将句子层面的对齐语料库转换为数据库后，借助用户图形界面，人们就能方便地利用 SQL 语言进行各种组合条件的全方面查询。元属性的限制条件通常是一个确定值或多个确定值的组合。确定值是二选一或多选一，例如翻译对的属性值是{英译汉，汉译英}取其一；时代属性值是{20 世纪前，20 世纪上半期，20 世纪下半期}取其一。对语言及翻译形式的检索则相对复杂。

当前针对翻译的查询方式主要有两种：1）语言学意义上的单语词或词与码的结合，如检索“研究”做动词和名词翻译成英语时采用什么样的译法。2）翻译研究意义上的源语言和目标语言之间的对应单元不同译法的检索，譬如中英文词语混合检索及出现与否的检索。

除了一些固定用法及成语（idiom）只能借助词典编码外，大多数翻译是按照源语言的组合方式，在目标语言中将与源语言对应的翻译单位重新组合而成，其中语法（如动宾关系）语义（如施事受事关系）等限制条件保持不变。以下以汉语动宾结构“克服 困难”为例介绍如何应用我们开发的检索工具，在英译汉及汉译英的翻译文本中检索相关的语言翻译现象，现在假定对文体、时代等其他任何元属性不做任何限制。语料规模为 200 万中英文字词。

如表 20.3 所示，利用“克服 困难”进行检索，可以检出汉语语句包含“克服困难”的英汉句对。如果想知道“克服 困难”中的“困难”翻译成“difficulty”的例子，可采用 A1 的检索串；想知道“克服 困难”中“困难”翻译成非“difficulty”的其他词语，采用 A2 的检索串；如果要检索汉语包含“克服 困难”，同时对应英文词分别翻为

"overcome difficulty"的实例，用 A11 串；检索"克服 困难"对应的英语词为非"overcome"和"difficulty"的实例，用 A12 串。检出的相应例句数参见频数栏，频数关系满足 A＝A1＋A2；A1＝A11＋A12。

表 20.2 利用不同检索字符串确定翻译对应

			检索串	频数	汉→英→汉
A			克服 困难	98	82:16
	A1		克服 困难 difficulty	86	76:10
		A11	克服 困难 overcome difficulty	70	68:2
		A12	克服 困难-overcome difficulty	16	8:8
	A2		克服 困难-difficulty	13	6:6

表 20.3 英语译词的选择

类型	"克服"		"困难"	
	不同词数	翻译实例分布	不同词数	翻译实例分布
汉译英	4	overcome (70), surmount/insurmountable (7/4), to do so (1)	4	difficulty (76), they/them (3), to do so (1), obstacles (2)
英译汉	10	overcome (4), Ø (4), transcend (1), contend with (1), fought (1), get out of (1), struggle with (1), straighten out (1), master (1), vanquish (1),	4	difficulty (10), Ø(3), hardship (2), obstacles (1)

使用 A 式检索，在本章实验所用语料实际得到 100 条记录。由于汉英句对都只经过词类赋码处理，而没有句法分析；当前检索对象只是在一个句对范围内汉语同时包含"克服 困难"的句子，不可避免带有垃圾。譬如汉英"克服投降，战胜困难"中"克服""困难"就在一句范围中，但不是检索所要的"克服困难"。经过人工排查过滤，最终满足要求的结果数为 98 例。

从表 20.3 中所示，检索汉语句子包含"克服 困难"的句对，绝大多数是汉译英文本，占 83%，少部分是英译汉文本，只占 17%。通过对这些检索出来的例句，我们至少能够发现如下现象：

20.4.1 汉译英的英文句子大多中规中矩

这些翻译句对的文本属性多数是政论文体。将"克服"翻译成英语时，主要有两种译法，分别是"overcome"(70 例，占 85.4%)，"surmount"(11 例，占 13.4%，包含四例"不可克服"译作"insurmountable"的句子)。经查询汉英词典，"克服"在汉

译英时基本上是取其高频对应词而成①。“困难”的翻译也基本如此。直译为“difficulty”有 76 例，占总数的 92.7%；剩余六例非 difficulty 的翻译实例中，有三例是采用代词“them”或“to”，在当前英语句子中找不到相应的“困难”译词，存在承前省略现象；另有两例翻译作“obstacle”，如“克服暂时的困难/surmount the temporary obstacles”“克服自己道路上的各种困难/surmount all the obstacles in their path”，形成 surmount 和 obstacle 的搭配。

利用“Google 翻译”来验证“克服困难”的英译②，其“翻译：汉语≫英语”栏目结果正是“to overcome difficulties”；而“字典：”栏目显示的是“surmount difficulties”。用“雅虎翻译”③来翻译“克服困难”，结果是“overcomes difficulties”。这与我们从汉译英语料中观察到的结果相一致。

按照汉语动宾关系的实现模式，根据词典释义词或词表对译词直接套用生成英语译文，是忠实原文的表现；同时间接反映出作为外语，中国译者的语言表达的丰富性还是比较欠缺。

20.4.2 英译汉的实例较少，用词丰富

从英语翻译成汉语，可理解成两个连续的过程：其一是根据英语形式理解其意思，这是个从形式到意义的语言分析（analysis）过程；其二是将该意思以汉语表述出来，这是个从意义到形式的汉语生成（generation）过程。由于有不同句式及词汇层面的不同表述方式，因此在反映“用坚强的意志和力量战胜（困难）”这一意义时，在汉语中分别有“战胜”、“克服”、“跨越”等多个不同的同义或近义词语可以用来表述。由于有更多的可选词语竞争，因此能够完全匹配“克服困难”这四个字的汉语译文自然就相对较少。

同时还要看到，在英语中，可以用来表达类似“用坚强的意志和力量战胜（困难）”意义的词或短语也有不少。譬如在 WordNet 中与该意义等值的“deal with successfully”的同义词集就有“overcome, subdue, surmount, master”等词，如果再扩展一下，加入近义或类义表示，那么可选的表达方式就很可观了。在英译汉的文本中，母语为英语的作者可以选用最贴切的表述方式，而不管采用词语或是短语形式实现。这使得英语表达更丰富，除了可以见诸词典的常规表示，还有更多基于上

---

① “克服”根据《远东汉英大辞典》（张芳杰主编，新华出版社/远东图书公司，1994）第一个义项为 to overcome (difficulties, etc)；《现代汉英词典》（本社词典编辑室，外语教学与研究出版社，1988）第一个义项为（战胜）surmount；overcome；conquer.

② http://translate.google.cn/translate_t?q=克服困难#zh-CN|en|克服困难

③ http://fanyi.cn.yahoo.com/translate_txt

下文的一次性用法。

在82个汉译英"克服"实例,翻译后的英语句子中只有三个不同词形翻译(surmount 与 insurmountable 视作同类),而只有16个英译汉的实例却有十个不同英语词形。"困难"的翻译在汉译英及英译汉都只有四种不同词形。由此看出,英译汉实例中的意义表述用语比较发散,而且有相当数量的短语形式,如 contend with, get out of, struggle with, straighten out 等。与汉译英的英语用词较集中,可预测性较强相比,英译汉中的英语表述更加自然地道,需要特别学习。

20.4.3 英译汉有增译现象

对"克服 困难"的汉译英文本基本上没有发现漏译的情况,这可能与其多数是政论文体有关。一般而言,对政论文体的翻译相对正式。譬如一些重要决议、领导人的汉语著作都是久经锤炼的经典,为保证译文的准确达意,对这些文本的翻译通常都没有删减。而在英译汉文本中"克服困难"发现有四处增译,它们都不是像汉译英那样的政论文本。具体实例见下:

(1) 即使当时医生要克服的困难比现在还要大得多,在他那坚持不懈的努力之下困难也是会退让的。Greater things than the Doctor had at that time to contend with, would have yielded before his persevering purpose.

(2) 要是她能够克服这一切困难到达爱伦身边,那就好了! If she could only win to Ellen!

(3) 这种努力非常顽强,大家全力以赴,痛苦不堪地克服困难,才把它压了下去。tried so hard that it was only kept under with great and painful difficulty;

(4) "那少年是克服了最大的困难才集中了全身的力量说出话来的,但是他的神色却起着可怕的强调作用。"It was with the greatest difficulty that the boy gathered bodily force to speak; But, his spirit spoke with a dreadful emphasis.

其中,例(1)英语原文是"thing(事情)",译成汉语时,根据与其构成动宾关系的动词短语 contend with 具有"对付困难或不愉快事情"[1]意义,而将"事情"具体化汉译为带有不好意味的"困难"。例(2)是完全增译,原文只是一句简短表示虚拟

---

[1] 根据 Longman Dictionary of Contemporary English 对 contend with 的解释为 to have to deal with something difficult or unpleasant。

语气的感叹句，经过翻译使“赢得爱伦”需要花费很大精力才能完成这种艰辛变得更为凸显。而例(3)、(4)增译了“克服”，原文中并没有突出“克服”，只是强调了“把它压下去”和“说出话来”具有一定的难度，间接表明要完成这些事情需要有一些特别的处理，因而汉译中加译了对困难的处理——“克服”这一原本没有的译文。

20.5 小结

在规划建设语料库的数据库结构并建成较大规模的汉英对应语料库之后，我们着手为广大用户提供简单方便的访问手段。通过良好的用户界面，生成合适的数据库查询语句，使得翻译研究和学习者无需掌握复杂的查询指令和更多的计算机操作技能，就能及时方便地利用经过语言学家和翻译学家精心整理加工的双语对应语料库。

当前检索系统能够进行语言学的特定词与词类查询，能够实现英汉汉英双向的翻译单位检出，能够满足元属性不同限制条件的混合检索。输出结果保存为HTML格式，查询词语高亮显示，方便对结果集的二次检索和对结果的显豁观察。

限于现有加工处理深度，当前采用检索词语句间共现作为例句检出的依据，在召回率(recall rate)得以保证的情况下，准确率(precision rate)将受到一定影响。虽然为人工翻译研究与教学服务的检索系统由于有人工参与，优先考虑召回率，可以考虑引入浅层语法分析，适当增加准确率的选项。另外系统现在还是采用查询词语出现与不出现的布尔查询模式，今后可以适当借鉴基于内容的信息检索向量空间等模式，实现近似句的查找。

第 21 章

语料库辅助翻译功能探讨

本章旨在比较机器翻译软件和翻译记忆软件对翻译的辅助作用，探讨语料库尤其是汉英对应语料库在英汉翻译中的使用方式、用途和独特优势。研究认为，双语对应语料库能提供针对性较强的参考译文，能够实施英-汉和汉-英双向检索，还能通过提供更大语境或者复合检索为翻译提供丰富的参考资源。可以让译者充分利用各种潜能高效高质地完成翻译任务。汉英双语对应语料库能够成为功能强大的机助翻译工具。

21.1 引言

在信息技术高度发达的今天，凡与翻译相关的各种活动都要使用电脑或网络资源，各种电子资源和工具（电子词典、翻译辅助软件和网络资源）正广泛应用于翻译活动。半个多世纪以前，人们对机器翻译（MT）寄予厚望，希望它能够借重信息技术提供的各种潜力成就高质量的翻译，但机器翻译至今（在英汉语之间）还未开发成真正可靠的翻译工具。自 20 世纪 90 年代以来，机辅翻译（CAT）工具蓬勃发展，有许多翻译记忆（TM）软件相继得到开发和应用。此外，一些在线双语对译检索系统也出现在互联网上，展现出良好的应用前景。

与此同时，语料库的发展和应用给机辅翻译注入了新的活力。近 20 年来，随着各种语料库的创建和应用，双语语料库正以其独特的方式进入人们的视野，其辅助功能日益凸显。本章以北京外国语大学汉英对应语料库（CEPC）为例对比语料库与其他翻译辅助工具在翻译实务上的用途，重点探讨 CEPC 的翻译辅助功能。

21.2 机辅翻译工具的类型和特点

信息技术发展到今天，联网计算机犹如一个工作间，可以从那里存贮和调用翻译所需的各种资源，其数量、类型、内容及获取速度远胜过其他传统资源。除了电子词典和搜索引擎外，目前常用的英汉机辅翻译工具软件还有四类：机器翻译软

件、翻译记忆软件、双语对译句搜索引擎、双语对应语料库。这些软件对翻译都有辅助作用,但辅助方式不同,作用各异,这是本章探讨问题。

21.2.1 机器翻译软件(machine translaiton)

机器翻译软件很多,常见的在线工具有 Babylon Pro、华建翻译、Google 翻译工具、Free Online Language Translator、Babel Fish、Yahoo、有道,等等。这些软件都能产出完整的译文,但译文质量较低。表现在以下几个方面:

第一,一次只提供一种译文。如果不同软件给出的译文不一致,译者难以做出选择。例如"垃圾邮件"既对应 spam(Google),又对应 junk mail(华建/Yahoo/Worldlingo)。"体制改革"里的"体制"对应 institutional(Babylon、Google)、system(华建/有道)和 organizational(Worldlingo、Yahoo)。究竟该选用哪个?软件本身无法提供更多的信息,这时机器翻译对译者助益不大。

第二,词库更新不够及时。一些较新的词语组合找不到合适的译文,如,"山寨(手)机"有多个应词:Mountain fastness machine(华建)、Village in a mountainous area machine(Yahoo)以及"山寨 machine"(Google)。而"群体事件"的翻译更是鱼龙混杂,有 groups events, colony incident, mass incidents, community event。没有其他信息做参照,面对这么多对应词,译者无所适从。(这一点从叙述上来说也属于"多个译文,难以取决"吧?)

第三,难以处理结构较复杂的汉语表达形式。比如,有的软件在翻译"踩踏事件"时只是根据原文将词机械拼凑起来的,加上冠词,如 trample the incident, steps on the event。译文合乎语法,但没有意义。再如,在处理"你真记性坏"这个句子时,出现下面两种译文:

a. you really bad memory. (Babylon/Google)
b. Your real memory is bad. (华建/Wordlingo/Yahoo)

"你真记性坏"这个句子的特点是:小句主语(记性)隶属于大句主语代表的事物(你)。句 b 显示,软件翻译能通过语义运算识别出"记性"和"你"之间的关系,但不能处理好副词"真"。而能处理副词"真"的 a 句却不能处理好"记性"和"你"的隶属关系。可见,机器翻译对于汉语句法关系的判断能力还是较弱,语义运算能力有限。

第四,语序调整能力差,对于汉语话题句的处理能力很低。比如,"无线电我是门外汉"是汉语常见的话题句,大句主语(无线电)前边隐含"对于、关于"的意思。英语没有这种句式,翻译时要做出添加或者调整。但现有的机器翻译软件均无法对此做出恰当的调整,于是产生下面的译文:

a. Radio I am a layman. (Babylon/Worldlingo/Yahoo)
b. Radio I layman. (华建)
c. I am a layman radio. (Google)

简言之,现有的翻译软件主要靠词对应实施翻译,内容更新不及时,语义运算能力差,句法识别和转换能力有限。机器翻译主要关注源语,而人工翻译的目标是理解目的语,所以前者译文自然难以准确。正如冯志伟(1996:424)所言,机器翻译软件对于一些复杂的句子的分析依然很困难,多义词和歧异问题尚未找到切实有效的解决办法。

21.2.2 翻译记忆软件(translation memory)

面对机器翻译难以逾越的语义瓶颈,美国政府曾责成ALPAC于1966年对这类系统进行评估,结果显示,机器翻译速度慢、效率低,比人工翻译贵一倍,所以,这类软件的开发得不偿失(Craciunescu 2004)。在沉寂了多年之后,从20世纪90年代开始,机器翻译领域发生了重大转向:从基于语法规则的翻译转变为基于翻译实例的翻译,翻译记忆的设计理念应时而生。人们对基于翻译实例的翻译记忆工具兴趣渐浓,可以想见,若能将由语料库转化而成的语言知识库同基于规则的机器翻译结合起来,机辅翻译前景光明(Bernardini et al. 2003:3)。

目前,不少翻译记忆软件已具备商业用途,如Déjà Vu和Trados,还有我国的雅信CAT。这类软件具有自动记忆和匹配功能,可根据译者需要自行建库,还可以有效地进行术语管理,确保科技翻译中术语一致。另外,此类软件能实施多语种双向互译。现以Trados为例考察它的检索功能。

Trados翻译记忆软件的检索功能和在线对译句检索程序工作机制相似,它能建立句库,可以帮助使用者有效地应对专业翻译。例如,从我们建立的Trados句对齐库中可以检索到"垃圾邮件"的两个对应项(见图21.1)。

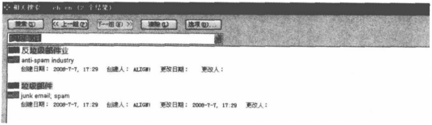


图21.1 Trados的检索结果

要像普通对译句检索那样提供丰富的对译,翻译记忆软件就需建立相当规模

的句库。但对于个体译者而言,限于时间、精力和经济条件,大规模建库不太现实。但不管怎么说,翻译记忆软件有助于提高翻译质量和效率,特别适合专业文献翻译。

21.2.3 在线英汉双语对译句子检索平台

汉英在线双语对译句子检索工具出现时间不长,近来比较活跃的有句酷、金山爱词霸和融合句酷及网络释义的灵格斯对译句搜索引擎。支持这些在线工具的句库已具备相当规模,且检索方便。此类工具的辅助能力表现在以下几个方面。

第一,一次能检索到例证的丰富。例如,"群体事件"但在灵格斯中有很多在句子中出现的对应项,有 mass incident/event/disturbance;group affair/event/accident;colony event,甚至还有 mass emergency。这显然要比机器翻译能提供更多的选择,只是对应项太多,也不容易做出选择。

第二,多源检索和呈现。现有的许多在线句对译搜索工具附有词典或百科搜索,有时有较高的参考价值。例如,爱词霸附带的"百科"搜索显示,"山寨机"的英文翻译是 emulational mobile phone。虽然我们还无法确认这样翻译是否合适,但至少在意思上说得通。

第三,有时也能检索到比较可靠的译文。如"踩踏事件"在灵格斯(句酷)和爱词霸百科中均发现对应项"stampede"。甚至还有对译句,如:

> Higher inflation has recently triggered popular reaction in Asia ranging from factory strikes in Vietnam to **stampedes** for discounted items at some Chinese supermarkets.
>
> 近来,通胀加剧已在亚洲引发普遍反应,例如从越南的工厂罢工,到中国一些超市促销引发的**踩踏事件**等。

第四,在句子层面上,译文的参照价值高。如:"我真记性坏。"就碰巧遇到非常相近的句子:

> My memory is really dreadful—manager's daughter?"(句酷、灵格斯)
> 我记性真坏——经理的小姐?"

有些译文提供的翻译方式还相当丰富,完全可以帮助我们产出高质量的译文。这是机器翻译软件难以企及的。如"无线电我是门外汉。"可以有很多值得参照的对译词和句子结构。

When it comes to mathematics, I'm completely at sea.（句酷、爱词霸）
说到数学，我完全是门外汉。
Where the law is concerned, I am only a layman.（句酷、灵格斯）
谈到法律，我不过是门外汉。
All that form of life was outside him.（灵格斯）
对于那种生活方式，他全然是个门外汉。

有些译文大同小异，甚至可以直接使用现成的翻译。如“你这种事做不得。”可以直接得到下面的译文：

You must never do this.（句酷）
这件事你做不得。
You must not do it.（句酷）
你不得做这件事。

尽管如此，对译句检索平台的辅助功能还是十分有限。首先，语境信息不够充分，只考察句子对译不足以应对各种复杂的情形。其次，现有对译句检索平台没有系统地区分源语和目的语，对译成分的选择比较盲目，难以保证译文地道。再次，对译句只是陈列多个译句，难以考察影响语言运用的其他因素（如语域、文体、搭配能力等）。

另外，现有的对译句检索多来自双语词典中的例句，语料来源不丰富，软件使用的词典有很多雷同。而且，受制于语料库规模，供检索的内容有限。例如，“垃圾邮件”在句酷和灵格斯上有两个对应项：junk mail 和 spam。该选用其中的哪个？我们心里还是没底。

21.3 双语对应语料库辅助翻译

现有的在线翻译辅助资源只是句对齐文本，没有更大的语境参照，也没有进行语法或其他功能标注。而且，现有的句对齐检索仅能显示非常笼统的元信息（*Metainformation*，如该句来自“词典”或来自“网友提供”），参考信息不充分。在检索功能上，翻译记忆工具虽能提供词精确和模糊检索，但不能进行词性检索，更不能检索到翻译方向、文体等重要信息。

要避免上述问题，双语对应语料库是一个很好选择。北京外国语大学的汉英对应语料库（CEPC）创建于 2004 年，该语料库语料规模还在不断扩充之中。其主

要特点是：1)语料规模大，目前的可检索语料达2000万字词，检索内容丰富。2)汉英双向，既有汉英翻译语料，又有英汉翻译语料。3)语料制作过程复杂，涉及句对齐，xml编码，有用的元信息和英汉语词性标注。4)分类细致，有题材、语域、翻译时间等详细的分类。5)检索方式多样，既有可分类的在线检索，又有较复杂的研究用检索。

对应语料库已广泛应用于翻译研究，但有关它在翻译实践上的应用研究还很少，有关使用者如何有效使用该类语料库的研究更少(Santos & Frankenberg-Garcia 2007)。所以，有必要介绍汉英对应语料库在翻译中可能的用途和应用方式。

21.3.1 参考译文针对性强

CEPC可以根据词汇使用的语域和语境检索对应词。例如，在“只要不动摇、不懈怠、不折腾”中“折腾”并不指具体的动作，而是一类社会现象。从其他资源中得到的多是具体的动作，如“turn from side to side, toss, do something again and again”。但在CEPC中，可以检索到更有针对性的例证。如：

(1) 如果搞资产阶级自由化，就是再来一次[折腾](汉英：非文学)

Bourgeois liberalization would plunge the country into turmoil once more.

(2) 因为中国不能再[折腾]，不能再动荡(汉英：非文学)

Because China could not afford any more disorder or unrest.

(3) 因为萨利已经开始瞎[折腾]了(英汉：文学)

For Sally was crumbling.

(4) 在自己的宅子里头瞎[折腾](英汉：文学)

defiling his own palace with it;

不难看出，作为一种社会行为的“折腾”在英语里的对译可以是plunge... into turmoil或者disorder。

21.3.2 双向检索例证

双向检索指可以根据不同的源语和目的语关系检索例证，好处是多向验证。例如，“我真记性坏”的英译可以从汉英和英汉两个方向寻找翻译对应项，这样有望检索到适切的对应词或对应结构。

(5) 我记性真坏——经理的小姐？(汉英)

My memory is really dreadful—manager's daughter?

(6) 文纨，你记性真好！(汉英)

What a good memory you have, Wen-wan!

You've neither sense nor memory, and I leave it to fancy where your mothers was that let you come to sea.(英汉)

(7) 你们既没有头脑也没有记性，我真不明白你们的母亲怎么会让你们出海。

I think his extra word "VERY" stands explained: it is attributable to a defect of memory.(英汉)

这多出来的一个字就已经说明问题：这是因为[记性]差了那么一点儿。

通过英汉翻译，我们可以确定 memory 对应"记性"；通过汉英翻译，我们可以确认"你真记性坏"可能使用 you really have [...]，或者 how adj. [...]，或者 what a... 结构形式，据此可形成下面的译文：You really have a bad memory. / How bad your memory is! /What a bad memory you have.

21.3.3 提供更地道的译文

根据(秦洪武、王克非 2004，2009)，英汉翻译时，译者会过度使用汉语里的某一表达式或者结构来翻译英语里的某类表达式。如过度使用汉语数量短语"这种、这一、这个"来翻译英语指示代词 this/such。实际上，"这种"在汉语里用得不多，且与之共现的词项有时有贬义。英译汉时过度使用这个短语是受英语影响造成的。要选择多样的表达，译者可以检索汉英翻译库中检索 this/such 的汉语对译文，这可能找到更符合汉语表达习惯的处理方式。

(8) 我们要抓住时机，现在是改革的最好时机

We must seize [this] highly opportune moment for reform

(9) 又是范博文的刻薄的声调

[This] was another cut from Fan Po-wen

(10) 这么，我们公司眼前既有事业好做，将来"东方大港"之类完成了的时候，我们的事业就更多了

In [this] way, we'll have something solid to get on with, and then one day, when the "Great Eastern Harbour" and so on have been completed, we'll be able to expand even more

(11) 这样做不会也不可能破坏社会主义经济

[This] cannot and will not undermine the socialist economy

由此反推，英语里好多使用指示(如 this，that，such)的地方在译成汉语时不需要任何指示语；而且，“这么”、“这样做”、“此”都可能对应 this。总之，译者通过观察这类语料反思自己的翻译行为(王克非、秦洪武 2007；秦洪武、王克非 2007)，避免在翻译时过度使用“这种”。

21.3.4 提供更多的语境信息

使用较大的语境有助于我们弄清一个词在具体语境下的用法，以及这些用法的语用含义。根据语境，第一个例证里的“外行”有自谦的口吻，而第二个例证里有调侃他人的意味。这些信息在翻译时很好的参照作用(见图 21.2)。

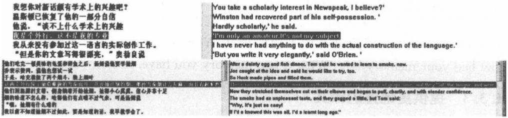


图 21.2 相关检索结果

CEPC 还可以使用通配符检索、词性检索，并提供检索词的搭配信息，其他检索工具都不能实施此类检索。这些检索能够从多个角度确定一个词的用法。

21.3.5 复合检索功能

复合检索功能即运用语料提供的元信息，使用多个检索限制条件提取译者需要的信息。CEPC 目前推出的在线检索程序(“中英双语在线”)能够实施此类检索。检索方式见图 21.3。

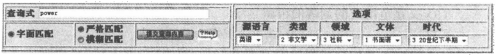


图 21.3 CEPC 检索方式

图 21.3 显示，在线检索系统可以执行严格匹配和模糊匹配，模糊匹配有助于许多结构式的检索。更重要的是，该检索系统较完整地实施了复合检索：如源语设定(英语或者汉语)、类型选择(有文学和非文学)、领域区分(文学、人文、社科、自

然)、文体分类，等等。例如，若设定了检索条件，可得到“与其……不如”的索引项(见图21.4)。



图21.4 “与其……不如”的索引项

检索结果显示，在20世纪上半期的英汉翻译里，对应英语 rather... than/more adj. ... than/not as adj. as 的很少使用“与其……不如”(一次)；而在20世纪后半期的英汉翻译中，则出现了五次。而汉语原创语言中这类结构没有出现。看来，现代汉语有更灵活的方式来表达“与其不如”的意思。这时可以非常方便地检索源语言为汉语的子库，我们可以检索到对应目的语英语 rather... than/more... than/not as adj. as 的汉语表达式。如：(not as adj. as 一般翻译也好像较少使用“与其不如”?)

(12) “赠人以车，不如赠人以言”

“A gift of a cart is not as good as a gift of words. ”;

(13) “跟车铺打印子，还不如给我一分利呢!”

“Better to pay me ten per cent than pay the shop installment. ”

(14) “我以为这一点的可能性很大，他马上就会看到独角戏不如搭班子好。”

“Not at all: I’m sure there’s every possibility that he’ll soon realize that it’s better to play a part in a big show than try to run a one-man show. ”

(15) 与其热枕头上翻来覆去，还是甲板上坐坐罢。

Rather than toss and turn on his warm pillow, Hung-chien would just as soon sit for a while longer on the deck.

(16) 他叹着气说：“唉，与其因为他有才能而被夺，我宁愿他是一个一无所长的没出息的人呵。”

He said with a sigh, “I would rather he be a mediocrity than a talent ed person and lose his life. ”

(17) 故此主张“与其忠实而使人看不懂，毋宁不很忠实而看得懂。”

Thus, they argue: "'Unintelligibility' resulting from faithfulness is worse than faithlessness that makes translation 'intelligible'."

很明显,"与其……不如"这么完整又有点机械的结构只有在英汉翻译里使用,汉语里还有更多的组合方式可供选择,这也是英汉翻译时译者应参考这些表达方式,以避免译文流于单调、机械(见图 21.5)。

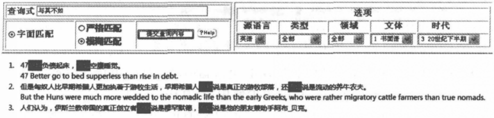


图 21.5 "与其……不如"的相关检索结果

以上检索功能的最大优点是:在语料库足够大且分类明晰的情况下,译者可以从中得到分类明确、针对性较强的参考译文,这些十分有用的功能其他翻译辅助工具都不具备。随着语库规模的扩大和检索功能的完善,CEPC 的翻译辅助功能会更加突出。

21.4 单语语料库辅助翻译

翻译时常常遇到难以把握的词汇运用(词汇选择、词汇搭配、句法关系),这时可以参考目的语原创语料库,因为原创单语语料库可以提供更丰富的语言使用信息。

前文说过,爱词霸"百科"搜索可以查到"山寨机"的英文对应表达 emulational mobile phone,而灵格斯的"网络释义"提供了"山寨"的对应词 copycatting。这两个词是否都合适呢? 我们在 COCA 中没有发现 emulational。emulation 的意思是仿真,它本身就是一种计算机技术,不是侵权仿制。但在 COCA 上检索到 copycat 的使用。如:

(1) Illegal sales over the phone and the Internet have been reported, and copycat products have already appeared.

Audiences might now expect to see copycat movies. "Once one movie hits, everybody usually tries to make a lot of them," says Adam Schroeder.

这两个例子里的 copycat products 和 copycat movies 可以分别译成“山寨产品”和“山寨版电影”。这样说来，“山寨机”译为“copycat mobile phone”该不会有问题。

再如，“垃圾邮件”有 junk mail 和 spam 两个对应项。COCA 的检索结果显示，spam 使用的频率远远高于 junk mail。而且，Yahoo 英文网站的邮件系统里就是用 spam 指垃圾邮件。这样，“垃圾邮件”的典型对应词应该是 spam。

面对“群体事件”令人眼花缭乱的对应项（mass incident/event/disturbance; group affair/event/accident; colony event, mass emergency）。我们进入 COCA 语料库尝试检索，发现，group event 常用，但意思是集体项目；只有“mass incident”被认可，尽管它还不是正式的用语；其他均无参考价值。百度和 Google 搜索常见的是 group event 和 colony event，但大都与 CNKI 知识搜索有关，是汉语学术文献的译法，同样不可取。以下是 COCA 中检索到的用法：

> (1) popular discontent is bubbling in China. Last year, there were 58 000 **"mass incidents"** across China, according to government statistics cited in the Chinese magazine Outlook.
>
> (2) Chinese government's own figures, last year there were more than 80 000 **"mass incidents"** throughout the country. It is commonly believed that a great number of these
>
> (3) erupt on a regular basis. The Ministry of Public Security reported 17 900 **"mass incidents"** in the first nine months of 2006. They are almost always snuffed out

另外，“踩踏事件”在灵格斯（句酷）和爱词霸中各发现一个对应项，即 stampede。在 COCA 中我们还发现了更多的对应项。如：

> Mayor Daley's top lawyer told a judge today that one of the owners of the Near South Side nightclub where 21 people died **in a stampede** should be thrown in

这说明，“踩踏事件”最合适的译文是 stampede，而不是其他词。机器翻译的不可靠性可见一斑。（以上没有指出 COCA 有其他更多的对应项，只是进一步验证了确实灵格斯（句酷）和爱词霸中存在“stampede”的使用。也不能以这一点证明机器翻译的不可靠性。

在实际翻译时，汉英对应语料库中的原创文本库本身就可以用作单语库，常用

词的组合和搭配可以在这类库中检索。如果对应库解决不了,就需要通过检索更大的单语库来确定用法。由此看来,单语和双语语料库在翻译实务中的用途是其他各类翻译辅助软件无法替代的。

21.5 小结

机器翻译在翻译辅助方面的功用有限;基于翻译实例的翻译记忆对翻译的辅助作用高于机器翻译,但由于检索功能过于单调,辅助作用还是有限。相比而言,双向双语对应语料库对翻译辅助作用更加突出,它可能提供的丰富的翻译对应信息其他工具无法提供。因此,在双语对应语料库这个操作平台上,译者可以充分利用它的各种潜能,高效高质地完成翻译任务。我们相信,随着制作和检索技术日趋完善,汉英对应语料库将成为功能强大的机助翻译工具。

附:本章涉及的主要在线检索平台网址

Babel Fish:http://babelfish.yahoo.com/

Babylon Pro:http://translation.babylon.com/

Free Online Language Translator: http://www.worldlingo.com/en/products_services/worldlingo_translator.html

Google 语言工具:http://translate.google.cn/translate_t?hl=zh-CN#

Yahoo 翻译:http://fanyi.yahoo.com.cn/translate_txt

北外中英双语在线(CEO):http://202.204.128.82/ceo/index1.html

华建翻译:http://www.hjtrans.com/

金山爱词霸:http://www.iciba.com/

句酷:http://www.jukuu.com/

有道翻译:http://fanyi.youdao.com/

第 22 章

专门对应语料库的研制

在评介国内外现有单语或双语旅游专门语料库的基础上，本章着重论述一种新型中英双语旅游语料库的研制，特别是其设计理念和操作程序，包括语料文本的数字化、语料的标注、语料的对齐和语料的篇头标注等。本章最后论述了旅游语料库在推动旅游翻译研究与教学的前景。

22.1 引言

过去 20 年来，基于语料库的翻译研究从无到有，发展迅速。目前，翻译研究中常用的语料库主要有译文语料库(translational corpus)、类比语料库(comparable corpora)和对应语料库(parallel corpus)三种。译文语料库收录译文，其研究目的在于揭示翻译语言本身独有特征(如 Mona Baker 建立的“翻译英语语料库”，即 TEC)。类比语料库主要是同一种语言的原生文本和翻译文本(译自某种外语)构建的语料库，它们之间无翻译对应关系，但在时代、体裁、主题等方面具有可比性，可用来研究翻译语言的特点。这类研究可参看 Laviosa(2002)。

相对于以上两种语料库，翻译研究中更为常用的是对应语料库，它收录原文与译文双语对照的文本，通常会按事先设定的标准(如以句或段为单位的方式)，对语料进行某一语言单位上的对齐(如句或段)以方便检索。对应语料库常用于考察原文中的某些语言现象如何在译文中得到反映，最终目的在于揭示翻译活动中隐性的规律。目前对应语料库在不同的国家和地区都有广泛的利用，比较知名的有挪威的“英语—挪威语对应语料库”、英国的“德语—英语文学文本对应语料库”以及我国由北京外国语大学中国外语教育研究中心研制的“汉英对应语料库”等。

纵观目前为翻译研究而建的上述三种语料库，多数是以文学文本为主，或以收纳百科文本(包括各种文学与非文学作品)为特色的通用型、综合型语料库，以及法律、财经、农林、医学等专科型语料库，是比较大型的双语语料库，而结合地域特点的或针对某一文类而专门研制的双语专门语料库较为少见。专门语料库指的是“关于特定主题文本的集合”，且这些文本“均由行内专家为不同的读者群所写”，这

些读者群可能是同行的专家，或是缺乏相关专业知识的大众，亦可能是学生等(Kübler 2003:29)。双语的专门语料库则收集包括原文与译文的专题文本。

迄今为止，专门为翻译研究而研制的应用型双语专门语料库并不多，且通常规模较小，从几万词、几十万词到百万词不等。小型的双语专门语料库由于其规模小、建造方便和针对性强等特点，常在术语翻译研究或培训特定领域(如专门法律翻译、特定财经翻译等)的译员中作为辅助工具，如 Kübler(2003)为培训软件指南译员而建立的小型英法计算机语料库。

双语专门语料库无论对翻译研究还是译员培训都有其潜在的价值。早在1991 年 Johansson 就提出，双语专门语料库可以帮助我们解决不少特定的问题，因而在各种语料库中，尤其值得提倡。正如 Olohan 所言，专门语料库不仅能帮我们研究与普通语料库中同时存在的一些普遍的语言现象和规律，还能考察那些仅属于这个类型文本的一些与众不同的特点(2004:20)。从语言运用角度而言，双语专门语料库(尤其是非文学专门语料库)能帮助译者理解和掌握同一领域中不同语言的常用术语、惯用表达式、文章语气及典型的语篇结构，避免在专业翻译中经常出现的语言“不自然”或“形似而神不似”的现象。但时至今日，无论国内国外，较具规模的双语专门语料库还不多见，如旅游方面就仅有一些小规模的语料库。

22.2 现有旅游专门语料库

迄今，国内外专门收录旅游文本的单语或双语语料库均不多见。纵览相关文献，我们发现国外的旅游专门语料库主要有以下三个，分别为芬兰、英国和日本学者研制。

第一，芬兰萨翁林纳翻译研究学院(Savonlinna School of Translation Studies)的英语教师 Michael Wilkinson 研制的合共 67 万词英语旅游文本语料库。该语料库主要供芬兰学生将芬兰旅游文本译至英语时参考，检验他们的译文用词是否地道。同时，学生可以利用该语料库来“确认、证实或摒弃主观的决策，或是获得搭配方面的用法”(2005:1)。

第二，英国埃塞克斯大学(University of Essex)的研究生 Carlota Alcantar(2007)研制的英语旅游文本语料库。该语料库总词数为 37 795，收集的旅游文本分别来自 1)政府旅游部门的介绍；2)酒店、餐馆和旅行社的宣传资料；3)语言培训学校相关的资料。语料库研制目的在于总结出英语旅游文本写作中最常用的 421 个词汇和短语，以帮助墨西哥纳亚里特州立大学旅游学院的教师有效地教授旅游常用词汇。

第三，日本大学(Nihon University)的 Kiyomi Chujo 等(2006)研制的“京都旅

游语料库”(Kyoto Tourism Corpus)。此单语语料库收集的均为日本京都市用英文写就的关于京都旅游介绍的文本共885篇，平均每篇47个词，收录了2786个不同类符(types)和42025形符(tokens)。库中所有的词语都用CLAW7标注软件消除屈折形式，还按词性作了标注。语料库旨在归纳出适合于日本旅游市场(尤其是京都地区)的几百个最常用的旅游英语词汇，以供在日本大学学习旅游英语课程的不同程度的学习者使用。

我国国内研制旅游专门语料库更为鲜见。冯志伟(2001)提及，中国科学院自动化所建立了一个旅游咨询口语对话语料库和一个旅馆预定口语对话语料库，但关于这个口语语料库的情况却一直不为人知。除此之外，国内尚不见有其他专门研发的旅游语料库(无论是单语还是双语)。

综上所述，国内外旅游专门语料库的研制还很不够，现有的几个有如下一些特点：一是规模小，除了Wilkinson的语料库有67万词外，其他两个语料库的规模均不超过五万词，并且形式单一，均为单语(英语)语料库。二是语料库的设计比较简单，选择语料的方法含糊不清，也没有说明其选择语料的标准，容易让人质疑其语料的代表性。三是相对于其他综合语料库而言，这些语料库对语料的处理都很粗糙，除日本大学的语料库外，其他的均未对语料做后期整理(包括对齐、标注、加篇头等)。四是这些专门语料库的用途都较单一，都是为旅游英语(尤其是词汇)的教学而研制。至于旅游文本在结构、修辞上的特点等深层次的问题完全没有探究，也没有考虑到旅游翻译教学与研究上的问题。本章介绍的新型双语旅游语料库希望能克服以上弱点，既可直接用于指导现实中的旅游翻译(实践意义)，亦可用于促进旅游翻译教学和研究(理论意义)。

22.3　新型双语旅游语料库的设计思路和特点

本章以下着重描述我们新近研制的新型双语旅游语料库(获香港理工大学资助)。这是国内第一个较大容量的英汉/汉英旅游语料库。我们主要介绍这个语料库的研制思路和方法、语料库的构成和特点，并讨论专门语料库的发展前景。

如Kennedy所言，无论何种语料库，其整体的设计和语料的汇集对于“基于该语料库的研究的信度和效度都具有十分重大的作用”(1998:60)。它还直接影响到语料库所设计的研究目的和教学作用能不能有效实现。就本旅游语料库而言，设计目的有三：

一是研究用途。包括旅游语言本身特点的研究及旅游翻译特点的研究。前者研究旅游语言在语篇、修辞、词汇等语言层面上与普通语言不同的特点；后者则研究在翻译过程中旅游翻译与普通翻译的不同特点。研究问题包括：旅游翻译在语

言特点上是否存在与普通翻译一样的共性？抑或有自己另外的特点？这些特点是与特定时期的翻译规范有关还是超越时间、时空限制？

二是翻译教学用途。具体来说，本语料库将为教授中英双语旅游翻译的教师提供真实的教学材料，并为设计数据驱动教学（data-driven learning，简称 DDL）提供真实和多样化的资源。DDL 这里指的是“在课堂上利用电子计算机生成的索引（concordances）来帮助学生发掘目标语模式（pattern）中的规律，并根据索引结果来研发种种学习活动及设计学习练习”（Johns and King 1991：iii）。就本旅游语料库而言，就是根据语料库比较和检索的结果，开发出各种 DDL 的教学手段来辅助旅游翻译的教学。而学生亦能从语料库中揣摩职业旅游翻译译员所用的翻译策略，从比较、思考中学习他们老到的方法与技巧。开发 DDL 教学手段往往需要更大规模的语料库，本库建成后会持续扩容，以满足 DDL 需求。

三是实用性。即能够为国内外从事涉外旅游的英汉语从业者提供方便的旅游翻译参考或用作自学材料，提高他们的业务水平。

为了实现上述目标，本语料库在设计时必须严谨，确保语料库的效度和信度。语料库设计时要考虑的因素很多，大体可分为整体设计和具体操作两个层面。下面分述这两个层面。

22.3.1 整体层面的考虑

从性质来看，本语料库属于专用（specialized）、同质的（homogeneous）语料库（黄昌宁，2002），只收集与旅游景点介绍相关的中、英文原生性的（spontaneously sourced texts）（Mason 2001：68）旅游文本及相关的中-英互译文本（在时机成熟后亦会扩展至收集与旅游相关的行业，如酒店、交通等介绍的文本）。这就局限了本语料库的文本采样不能像异质语料库般采用随遇性的文本采样原则，而只能是采取目的性取样标准（purposeful sampling）。为了确保采集到的文本在内容上与语料库研制目的相关，在语言上符合标准，本语料建设中除了会实行多个专家把关制之外，还会推出同侪报告（peer debriefing）及研究成员的检视（member check）等上述方法来核实文本的适合性和代表性。同时，对于某些介绍比较多的旅游景点文本材料，在选取这些材料入库时我们亦会附加以三角验证法（triangulation），多渠道、多方式地反复核对入库文本（Li 2004）。以上这些确保文本数据有效性（validity）的做法均是传统的双语旅游语料库研制中所鲜见的。

下面我们就本语料库在构成、大小、语料的代表性及语料的选择等具体做法作进一步介绍。

22.3.1.1 语料库的构成和大小

从整体上看，本语料库由一个双语旅游翻译对应语料库（parallel corpus，简称

PC)与一个双语旅游翻译类比语料库(comparable corpora,简称CC)组成。在研究的第一阶段,这两个语料库暂定为各100万字/词(为统计方便,汉语部分按字数计算,英文部分按词数计算),目前各个语料库已经完成三分之二。按照我们的设计,PC与CC均为动态语料库(dynamic corpora),即在将来会按每两年一次的频率,把符合条件的文本补充至各库之中,并不断提高加工的深度。与那些只是选择特定时间范围内部分语言现象的静态语料库相比,本语料库的动态设计能够反映旅游用语及旅游翻译的时代变迁和语言风格的变化。动态设计的思路会让本旅游语料库的规模逐渐扩大,同时也为研究旅游翻译和旅游语言提供了纵向研究的可能。在现阶段,PC与CC合并起来有200万字词。语料库是否适用的主要因素并非全在于其规模的大或小,而主要在于"适用于语料库的文本是否容易获得,或是取决于其他因素,诸如文本是否需要手动标注等"(Hunston 2002:26),以及语料库能不能实现研制目的(如旅游翻译的教学)和解决预定的一些研究问题(如旅游语言的特点)。对于本研究而言,各100万字词的PC和CC,在语料收集和分析上较易做到,且在现阶段为达到这个目标,在财力和在人力的资源上尚可操控。

PC是一个双向的汉英/英汉旅游翻译语料库,先收录香港地区的英译汉和汉译英的旅游翻译文本,并逐步将收集的范围扩大到大陆、台湾以及全球华人地区的旅游文本。这些文本包括以网页形式的电子文本和以书面形式的文字文本(包括书籍、宣传册、旅游区内张贴的介绍等)。

CC主要收录香港地区为主的两岸三地非翻译(non-translation)的汉语和英文旅游原生性文本。英文旅游文本将涵盖英、美、澳等主要以英语为母语的国家(国别在语料中加以标识)。与PC一样,CC收录的文本范围同样包括来自电子、书面以及其他媒介形式的文本。在构建CC时,我们注重让汉语文本库与英文文本库在以下几个主要参数上具有一致性,以确保收录在CC中的中英文文本具有最大的可比性,如文本均全文收录、长度大致相近、文本创作的时间跨度均为过去十年、文本的作者尽可能多样等(Laviosa 1997)。这些参数的一致性至少能确保CC在设计上避免由于中英文本语料的不均衡而造成研究中出现错误的或有偏差的结论。

PC与CC相结合的设计考量在于让这两种语料库的特点互补。Johansson(1998:6)认为,PC研究中出现的一些不容易解答的问题,需以CC为比较参照系才能得到让人信服的答案。这些问题包括:译文在多大程度上体现普通语言的运用? 译文在多大程度上受到原文的影响? 以及译文具有什么样的总体特征? 等等。正因为PC与CC结合能在语料库翻译研究中发挥更大的增效作用,近来有不少翻译学者都提倡把PC中的一些发现运用在CC的背景框架中解释,充分发挥语料库翻译研究的定性与定量研究相结合的优势(Kenny 2006)。除此之外,本语料

库里具有双向性的PC与CC结合还能提供更广阔、更丰富的语言对比潜能。如图22.1所示,C表示本旅游语料库的组成部分。

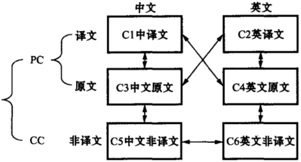


图 22.1 旅游语料库结构图①

从图22.1可见,若以库为单位,本语料库内可以作子库之间的比较,至少有以下几种可能性:1)可以是PC研究中常见的原文与译文的比较(C1:C4;C2:C3);2)可以作同一种语言之间译文与非译文的比较(C1:C3;C5:C3;C2:C4;C6:C4);3)亦可以是不同语言之间(这里即汉语与英文)非译文之间的比较(C5:C6)。同时,C5和C6还能起到参考语料库的作用,为C1:C4和C2:C3时提供尽可能多的参照系。本旅游语料库的这种设计方式为最大限度地有效利用语料库创造了条件。

PC与CC选取的旅游文本主要包括:1)自然景观、人文景观的介绍;2)观光小册子;3)酒店、旅馆、餐饮业及娱乐场所和设施的介绍;4)博物馆(包括展品)的说明和介绍。将来还可考虑收集与旅游业相关的其他行业的文本,如与交通、导游、票务、旅行社业务(包括旅游行程设计)、货币兑换、投诉处理、旅游安全、旅游管理部门的宣传和公示语等相关的文本。

22.3.1.2 语料的选取和代表性问题

香港旅游业兴旺,市场上充斥着大量与旅游相关的宣传页和小册子。数量多,质量难免良莠不齐,这就更凸现了语料选取过程的重要性。Kennedy特别指出语料的代表性问题——即所选语料能否代表这个领域中的大多数文本——是语料库建设必须考虑的一个重要问题。他认为,选择语料时必须注意的问题包括:语料的文学价值、社会上的影响、所面向的读者群体、文本地区性和时间因素等等(1998:62-66)。由于Kennedy针对的是单语普通语料库(尤其是文学语料库)的建设,他所提及的上述要素并非全都适用于本语料库。但就本语料库而言,在面对大量有关旅游的文本时,"研究者需要考虑一系列的问题,确定一系列的标准,帮助作出选

① 本图部分改编自Johansson和Hofland(1994:26)。

择和排除某些文本的决定”(Olohan 2004:46)。

借用 Maia(2003:45)的做法,本旅游语料库订出以下几条标准:1)据社会语域标准(如分为景点介绍、观光介绍、旅游配套设施介绍等若干类),将分层抽样、均衡抽样和比例抽样结合使用,使入库语料具有一定代表性;2)所有文本,若含任何语法、用词或印刷上的错误,均不收录,或略加更正后收录;3)就翻译文本而言,任何太过贴近原文句法结构以致僵硬的译文不予收录;4)在最后阶段,部分文本请精通中英双语的专家和翻译教师审阅其语言和用法上的可接受性。由于本旅游语料库除了用于语言研究外,还会用于教学,上述严谨选材除了尽量使语料具有代表性外,还要确保文本有一定的质量,这样才不会对旅游翻译教学造成反效果。

可能有人认为,语料若不是全部用随机方式而任意抽取,反而不利于呈现旅游文本的“最自然的状态”,因而语料库也并非所有香港旅游文本的最真实反映,而是带有很大的主观色彩。对此,我们认为:1)对语料进行适当选择是当今绝大部分语料库的通用做法,亦是一种发展趋势(如 BNC 就对不少语料进行了剪裁),只不过很多语料库建设者未将选材标准明确列出而已。未经挑选的大量原始语料的集合只会让语料库内容参差不齐,水平不一。由这些材料集合而成的语料库虽然最少受到研究者的“干扰”(即研究者主观性的影响),但它却不利于开展研究或教学,而后者恰恰才是语料库建设的目的;2)所有语料库都不可能做到完全客观,不受研究者主观性的影响。正如 Crisafulli 所言,“语料库从开始设计如就是一种(设计者)诠释的表现”(2002:32)。不少学者亦指出,语料库的设计以及语料的收集(包括要容纳哪些主要信息和次要信息等),都要根据研制语料库之前所设定的研究目的和研究问题而定(Olohan 2004:42),只要语料库的选材能够达到预定的目的,语料库的设计和语料的组成能与研究目的相符,能让学者进行相应的研究,这就是一个好的语料选取过程。

22.3.2 具体操作层面的考虑

语料库具体操作层面的考虑因素包括语料文本的数字化、语料的标注、语料的对齐、语料的篇头标注等实际问题。

第一,语料文本的数字化。目前,语料库所收录的文本绝大部分来源于以下三种途径:网上的电子文本、以纸质形式出现的印刷品、书籍或宣传页和拍摄下来的以建筑物形式树立的有关旅游景区的介绍。所有的这些材料都需转为 txt 文档备用。

不少印刷品都带有插图、照片、标志等 Nord 称之为“非言语因素”(non-verbal elements)(2005:118)以及用不同字体大小、形式、下划线等来模仿语调、语气和停顿功能的“超音段特征”(suprasegmental features),它们都构成了旅游文宣中不可

缺少的一部分。但在转换成语料库电子文本后，无论是“非言语因素”和“超音段特征”都得去除，以方便标注和以后的检索，因此，我们对于具有这些特征的文本就在其电子文本的“篇头”上注明，方便研究者和学习者如有需要时去打开相应的扫描文档来查找，同时亦会在文本中出现这些“超音段特征”的地方以赋码<SUS>做人工标注，让使用者能够通过计算机快速查找。以相片出现的旅游介绍则需要手工输入，之后亦存为 txt 文档备用。

第二，语料的标注。在本章，标注指对语料成分(通常是词性、词类)的划分，以方便进一步检索和研究，这种标注方式称之为 POS 标注法(part-of-speech tagging)。对于英文语料，一般的 POS 标注法就是参考 BNC 对语料词性和词类的划分和缩写方法(如 NN1 表示单数名词、NN2 表示复数名词、NP0 表示专有名词等，参看 BNC 网站)，先用软件(如 CLAWS 标注系统)进行机器标注。由于机器标注都不能达到百分百的标注准确率(最好的 CLAWS 标注系统也只能达到 96%左右，见 Kennedy 1998:218)，因此机器标注完成后还要进行人工修正。汉语亦有类似标注词性和词类的系统，但其准确性不如 CLAWS，且汉语词性的区分不如英文的清晰，因而标注汉语语料人工后期修正的时间要花费更多。

本语料库除了按照传统的 POS 标注外，还根据旅游文本中的一个突出的修辞特点，在语料上采用相应的修辞手段标注。研究员在研读文本时，分别会根据文本在语音、词汇和语法上的修辞特点，在语料中标出所使用的修辞或比喻手法。这种标注方法由于没有现成的软件可以做到，一切只能人工进行，所以费时甚巨。但一旦标注完成，它将为我们揭示英汉旅游文本在修辞用法上的不同以至于文本背后两种民族的审美观念的不同提供有力的佐证，这对比较修辞学、比较美学研究等都大有裨益。

针对旅游文本及旅游翻译的特点，本语料库亦会使用一些 CLAWS7 的码集中没有的自主创设码来对语料进行标注。这些自主创设码主要有两类：一类是用于标注上述提及的“超音段特征”，以创设码<SUS>表示；另一类是各种类型的修辞手段(rhetorical devices)，在创设码上全以<RHE>开头，后面会加以一个或两个字母来表明语料中所用的具体修辞手法，如隐喻以<RHEM>表示，头韵以<RHEA>表示，排比以<RHEP>表示等。因为 CLAWS7 软件只能标注词类，上述标注因而不能采用机器进行，而只手工赋码。创作这些码集的主要原因是因为它们所代表的语言特点在旅游文本体裁中非常显著，且作用重大，往往决定旅游文本的呼唤功能(vocative function)是否能以实现，因而有必要在语料库中将这些特点明示。

另外，标注还涉及 CLAWS7 的码集与自主创设码的兼容问题。因为作者并非 CLAWS7 的开发人员，亦无权将自创的码集加入至 CLAWS7 的码集之中，从而达

到以 CLAWS7 软件统一以计算机附码的目的。事实上，即便作者有权将自主创设码加入 CLAWS7 的码集，但对于作者自主创设码的标注（如修辞手段的标注），主要是基于对整个句子的认知理解，以目前人工智能的发展程度来看，计算机亦不能对此附码。为了与 CLAWS7 的码集相兼容，我们采取的手法是：在所有的语料用 CLAWS7 的码集标注后，再用手工形式对处理过的语料加注自主创设码。手工赋码费时，但准确性高。我们力求在手工赋码时做到标注格式与 CLAWS7 的码集格式完全一致，以方便在检索语料时用软件统一提取。

第三，语料的对齐。PC 里的中、英文语料都先用 ParaConc 软件以句为单位粗略对齐，然后再人工检查，以方便用 concordancer 检索。汉语语料还会用中科院计算所开发的“汉语词汇分析系统”（ICTCLAS）进行切分。旅游文本的翻译很多时候并非完全遵守句对句的原则，摘译、扩译或省译情况还比较多见。软件在某些地方找不到一对一的对应，就容易发生错配。我们设想将来在依照此新型双语旅游语料库的基础上，根据词频来自主开发出中、英文旅游文本中常用的双语词汇（bilingual lexicon），这样就方便语料库扩充和更新时，把这些双语词汇用作语料句与句对齐的另一附加条件，提高对齐的准确率。但无论如何，机器对齐之后，均需要人工检查作进一步的确定。对于摘译、扩译或是省译的现象，运用 concordancer 检索相关的句子时，可以在检索时设定 concordancer 同时出现相对应的句子的相邻的几个句子，提供更为详尽的语境，这样就能较为准确的知道是那一部分原文受到改动。语料对齐后，会将文本转换成 XML 格式，这样就能让语料库以多种语言格式编码，方便传播和检索。

CC 中汉语旅游语料的词汇切分与赋码亦是采用的是由中科院计算所的 ICTCLAS 软件。该系统分词精度高，分析速度佳，唯一的不足就是对词性的标注仍不够精准。因此，经过 ICTCLAS 分词后的语料仍需大量的手工校对，以提高汉语语料的词性标注正确率。

第四，语料的篇头标注。篇头提供所收语料的进一步信息。根据不同研制目的，不同语料库纳入篇头的内容亦不同。本语料库里，篇头内容包括篇名、作者名、作者背景、译者名、译者背景、年代、出版信息（书籍、电子或图片）、文本字数、有无“非言语因素”及“超音段特征”等。

从以上介绍可以看出，与传统的旅游语料库相比，本语料库的“新”主要体现在以下几点：1）从语料库的结构成分来看，本语料库一反传统旅游语料库均由 PC 单轨组成的特点，改为由 PC 与 CC 联合双轨组成。正如图 22.1 所示，这种组合形式能提够更为丰富的各子库之间比较的可能性，使研究者能更全面地考察旅游文本的体裁惯例，更深入地了解旅游翻译与其他类型的翻译（如文学翻译、法律翻译等）在词汇、句法和修辞等多个语言层面上呈现出来的不同特点；2）从语料库的词容来

看，本语料库中的PC与CC加起来共200万字，这亦是目前世界上最大的双语旅游语料库。更为重要的是，本语料库建成之后，每二年将更新一次语料库的内容，使语料库能反映旅游语言的最新变化。可以设想，在较长的一段时间之后，语料库这种动态的、增进式(incremental)的特点将会使研究者能够对旅游语言和旅游翻译的特点进行历时或共时的研究，从而丰富目前双语旅游语料库研究中较为单一的研究范式；3)从语料的收集上，本语料库的语料选材严谨，特别是针对双语旅游语料良莠不齐的普遍现象，提出双语语料均需由专家认定后才能入库。这就能避免目前某些双语旅游语料库中出现的部分收录译文错误百出的毛病，从而在保证本语料库的语料具有代表性的同时，亦不会误导用户；4)从语料库的运用前景上，传统旅游语料库作用相当单一，几乎都是为旅游英语教学，尤其是词汇教学而设；而本语料库的用途则广泛得多：它除了可以采用DDL形式，开展多层次、多类型的旅游英语教学与中英-英中旅游翻译教学之外，亦可以用于语言研究及供旅游从业人员使用，这样就使双语旅游语料库的价值得到更大的体现，亦能给其他专门语料库的研制和发展提供宝贵的经验。

22.4 小结

上述以旅游语料库为例讨论专门双语对应语料库的研制问题。这只是研制的一部分工作，还需不断扩充语料，完善标注体系，并根据研究和教学目的，进行新的语义和句法标注，最大限度地发挥语料库的潜能。在教学方面，将根据语料特点，开发出一系列可供教师在课堂上使用以及学生随时使用的教学材料，为开展DDL旅游翻译教学提供直观和真实的语料基础。另外还可以参照这个专门语料库，建立学习者旅游翻译语料库，收集大学生旅游翻译作业的电子文本，把它与其他两个语料库相对照，找出学生常犯的错误，寻求产生这些错误的文本内及文本外的原因。

专门双语对应语料库应用前景相当广阔。以此库为例，在词汇层面，我们可以比较汉英旅游语篇的词汇色彩和风格(如是正式、幽默、低调、口语化还是其他)，或是通过词频来统计出在特定旅游场景中最常用的词汇，作为编纂不同程度的旅游教材的基础，或是考察特定的形容词(如"美丽的")在译文中对应的词汇，从而挖掘旅游翻译过程中的审美因素等。在句子层面，可以考察汉英旅游语篇在句子结构和句子复杂程度的偏好，或是研究句子选择主动或被动表达式的深层原因，或是了解句子在不同的文化场合所传达的不同的语用效果，所能激发起读者不同的文化联想等；在语篇层面，可以比较汉英旅游语篇在衔接手段上的差别，或是归纳出它们在语步结构和体裁结构潜势上的不同特点，从而考察它们在社会认知根源上的

异同;在修辞层面,还可以比较汉英旅游语篇最常用的修辞手法,考察读者对这些手法的接受和反应,或是考察汉语旅游语篇中用到的特定的修辞手法在英文中的习惯处理方法,并了解它们在译文读者大脑中构成的或主观、或客观、或夸张的印象等。这些都是专门语料库研制后可见的研究课题和应用前景。

第 23 章

对应语料库的加工处理

本章围绕建设一个服务于翻译研究和教学的语料库平台，提出了汉英对应语料库的建库及检索分步处理的系统架构，完成了集句对齐与校对于一体的处理界面，提出了对既往收集语料的辅助校对策略方案。还介绍了对对齐后的双语句子如何进一步标注加工与校对，并转化为关系型数据库形式供检索调用。

23.1 引言

计算机和网络的发展使人们能够便捷地获取电子文本。海量的语言材料又进一步催生语料库的建设。当前语料库数量不断增加、加工处理技术不断更新，一些深加工语料库不仅进行词性赋码，还标注句法关系等(冯志伟 2007)。这些都已成为语言研究及应用的宝贵财富。由于有单语语料库的开发技术和机器翻译对大规模翻译语料的现实需求，更多的双语/多语语料库的研发已经提到学界的议事日程。

双语语料库可分为对应语料库、类比语料库和翻译语料库。一般认为，具有翻译关系并且相应文本之间具有对齐关系的对应语料库对翻译的教学与研究有效性更好。近年来，在教育部人文社科重点基地重大项目和国家社科基金项目的支持下，我们规划设计并已初步建成一个 2 000 万字词的“汉英对应语料库”(王克非 2004a；王克非 2004b；王克非 2004c)，又在此基础上进一步建设服务于翻译研究与翻译教学的平台(柯飞 2002；王克非等 2007)。为了使这一宝贵的语料资源能够尽快为广大翻译教师和研究者以及学习者服务，更好地接受各界反馈以利于语料库的发展完善，2008 年底，北京外国语大学中国外语教育研究中心开设的“中国外语教研网”[①]上发布了包含 200 万字词的“中英双语在线(CEO)”语料库及其检索系统。通过该检索人口，用户能够较方便地利用已建成的双语对应语料库中的部分测试资源。

本章主要介绍用于翻译教学与翻译研究的汉英对应语料库系统的设计架构，

① http://www.fleric.org.cn/ceo/

包括离线(offline)的课题组内部的语料对齐加工及校对，语料库的数据库转换等中间环节。

23.2 对应语料库

总体来看，建设双语语料库大部分是从统计机器翻译等自然语言处理应用项目中发展而来。例如可以对英法双语平行对应语料库通过训练，获取源语的语言模型 P(S)和源语向目标语转化的翻译模型 P(T|S)等参数估计，并将其用于统计机器翻译(Brown 1993)。基于实例的则更是从双语对应的实例库中找出与待译部分最相似的翻译实例(通常是句子)，然后对实例的译文通过串替换，串删除以及串增加等一系列变形操作实现翻译(Nagao 1984)。现在越来越多的自然语言处理项目都离不开各类语料库，为此，美国语言数据联盟(Language Data Consortium, LDC)[①]和欧洲语言资源协会(European Language Resource Association, ELRA)[②]已经成为协调和推动双语/多语语料库发展的重要组织。

对应语料库的一些成功应用使得语料库建设在世界各地蓬勃发展起来。例如利用加拿大议会辩论中的英法语文文本建成的 Hansard 英法平行语料库，IBM 公司运用统计方法以及各种对齐技术开发的 Candide 的统计机器翻译系统，与当时凝聚大量语言学家和语言工程师心血而研制的基于规则的机器翻译系统 Systran 相比，水平相差无几，从而推动了语言工程领域经验主义思潮的进一步发展(Berge 1994)。自此之后，建造各类对应语料库成为一种基础工作。国际上最近一个大项目是搜集欧洲议会从 1996 年到 2006 年 11 种主要语言各约 100 万句子构成的对应语料库，开发者宣称其目标就是为统计机器翻译服务(Koehn 2005)。

与此同时，我国语言工程领域的双语对应语料库发展也得到长足进步。一些主要的机器翻译研制单位，像北京大学计算语言所、中科院软件所、计算所及自动化所、哈尔滨工业大学、东北大学以及厦门大学等都构造了规模大小不等、标注层次不一的英汉双语语料库，其目的是服务于各自统计机器翻译系统的模型参数的训练要求。鉴于当前自然语言处理还是以句子为单位，因此大部分双语语料是以句对(sentence pair)形式出现。

应该指出，服务于机器翻译等项目的双语句对语料与通常供翻译教学与研究所用的双语语料库有很大不同，特别是在以下几个方面：

第一，研制目的不同。用于统计机器翻译的双语语料库更多地重视数量，因为

① http://www.ldc.upenn.edu/

② http://www.elra.info/

没有一定量的语言事实材料，在估算概率模型参数时会产生数据稀疏（data sparse）的情况，进而形成统计偏倚，造成翻译错误；供人们教学和研究使用的双语语料库在保证量的前提下更多地看重质，因此人工对语料的介入更多，如对语料进行机器加工之后，需要进行更多的校对处理，以保证语料库各类标注的质量。

第二，技术方法不同。统计机器翻译是以算法技术为追求目的，通过提取模型参数，选用统计算法，完成机器的自动处理。如何实现不同层次语言单位之间的对齐、如何度量待翻译语言单位与现有翻译单位的相似性，是统计机器翻译和基于实例的机器翻译重点考虑的技术路线；而供人使用的双语语料库在方便检索翻译实例的前提下，更注重发挥人的主观能动性，发现翻译规律。

第三，属性标注不同。供统计机器翻译用的句对语料注重源语言和目标语言之间的对应性，希望能对给定某种语言单位，通过对应语料库直接转换成目标语的相应译文，因此更注重语言的对译性；供人使用的双语语料库除此之外，还对各类其他属性要求更多，以方便进行不同类别的研究。譬如对年代属性进行历时分析，对源语言和目标语言属性进行英译汉或汉译英的分析。

用来训练和检验机器翻译系统句子翻译质量的海量句对，可以很方便地通过Web方式获得（叶莎妮等 2008）。一些 IT 公司在搜集此类语料后，还开发了免费双语句对检索工具，如爱词霸英语句库①、句酷②等，采用搜索优化技术，能够使用户很方便地在海量句对中定位检索句子。但遗憾的是其服务对象为普通用户，提供的检索界面比较简单，缺少对句子的一些元属性，比如翻译方向是英译汉还是中译英，句子所在篇章文本属性等，对翻译研究和学习者的帮助有限。

从语言学及翻译学研究出发，也有许多建设双语/多语语料库的典范，如英国曼彻斯特大学科技学院 1995 年创建的世界上第一个翻译语料库③，收集从各国语言翻译成英语的文本，规模以千万词计，并且语料采样建库明确。涉及英汉双语的语料库主要是旅居国外的中国学者和我国本土学者为主创建，如兰卡斯特大学的巴别（Babel）英汉平行语料库，南京大学、厦门大学等也有学者构建了不同规模的对应语料库（更多语料库详情可参阅“语料天涯”④）。

我们的“汉英对应语料库”建设，在兼顾机器翻译等应用的同时，更多的是希望给予翻译研究与教学更多的帮助，因而更倾向于为人服务而不是统计机器翻译的替换操作。

---

① http://dj.iciba.com

② http://www.jukuu.com

③ http://www.llc.manchester.ac.uk/ctis/research/english-corpus/

④ http://www.fleric.org.cn/sweccl/Corpus/

23.3 双语对应语料库平台的系统架构

经过项目组研发人员的多年努力，北外“汉英对应语料库”文字规模当前实际已经达到3000万字词，其中精细到词性赋码层次的句对齐(alignment)语料有30万句对(sentence pair)，基本能够满足日常翻译教学及研究。目前，有关更多语料的搜集、整理、加工、校对等各项工作还在有条不紊地进行之中。

整个系统架构如图23.1所示，逻辑上可分为开发阶段的内部离线的加工建库和对外开放应用阶段的外部在线的检索服务两个部分。物理上，这两个阶段并非截然分开的，建库一直都在进行，内部不断有新的加工完善的语料加入；同时还根据外部检索接口用户反馈内容对语料资源进行订正。外部接口界面设计也接受用于易用性的改进。

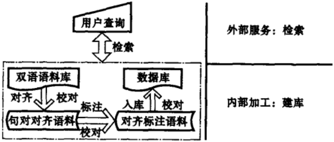


图23.1 汉英对应语料库架构设计

加工处理是系统研发人员将篇章对齐的汉英原始语料文本实施句子层面的对齐，将对齐后的双语句子分别做进一步的加工(如词性赋码、句法标注)；再根据语料库规划之初设定的标注属性(如标引源语言及目标语言、语体、领域等区别性属性)和检索原则(如句子接受模糊查询；接受不同属性组合的多条件复合查询等)，将加工好的句对转换成关系型数据库格式，供上线用户检索使用这一过程。这是一个长期细致的语言资源建设工程，不仅句对数量多、而且要求标注质量高。

检索服务建立在已创建的数据库基础上，是一个人机交互实时响应的过程，系统通过一个人机交互界面获取用户输入的查询条件，并对后台数据库存储的语句进行检索，并返回相应结果。

虽然当前自然语言处理技术，像双语句对齐、词性赋码及句法分析等都比以前有了长足的进步，但依然不能完全依赖这些技术手段，实现全自动无人值守的语料加工转化。在语料对齐建库、加工标注以及校对转换过程中，人工干预必不可少，尤其有经验的语言学家的指导作用更是不能忽视。为帮助不熟悉计算机操作的语

言学专业人员能够更方便快捷地参与工作，一个良好的用户人机界面必不可少。我们在整个语料建库加工处理过程中，设计了一整套辅助校核的工作界面，使得通常人工处理费时经年且繁杂易错的过程变得更为简单方便，准确性也得到保证。

建设一个服务翻译教学与研究的应用平台，需要经过一系列步骤：首先完成语料收集整理（包括纸版本语料的电子化、不同格式文件的纯文本化）、语料对齐的校对处理（句对齐），语料加工标注的校对，及将校核完毕的数据转存入数据库。要对这个平台加以利用则还需要提供一个用户访问获取该资源的检索接口设计。本章主要介绍内部加工的建库部分，有关用户如何检索利用该语料库进行翻译教学与研究，限于篇幅，另文撰述。

23.4 语料对齐加工、校对及入库

双语对应语料库的质量主要取决于语料对齐的质量和语料各类标注的质量。语料对齐可以分为不同层次，如篇章级、段落级、句子级以及短语、词语级。一般来说，层次越往上，重复率越低，意义也较确定。层次越往下，重复率越高，意义更多样。到词语层面，基本上就可以看作一个词库。在建设服务于翻译教学与研究的双语平台项目中，我们主要开发了双语语料的句对齐及校对、人工校对语料的形式再校对以及语料标注属性校对等辅助工具，实现语料及属性的形式和内容的校核，保证语料库质量，并在此基础上，将其转化为数据库，以便更好地利用数据库系统对语料进行全面管理和检索。

23.4.1 中英双语语料的句对齐及校对

最初搜集未经加工的生语料原文及对应翻译文本以纯文本（plain text）形式分开保存。原文文本及其翻译文本在命名设计上拥有相同且唯一的文件名，文件后缀冠以 cn，en 等区分中英文文件。因其名称相同，才能保证原文和译文是对应的；因其唯一，才能区分不同文件。实际处理时，只需定位找到具有相同文件名的两个文件，便可确定该文本的双语对应关系。

根据对前期整理的语料库统计（王克非 2004a），句子不失为翻译的一个主要转换单位。特别是对非文学类英汉双向翻译及文学类英译汉这三类翻译，1∶1的句对比例均达到 80%以上。即便文学类汉译英也有近 2/3 的句对比例。这一发现给予的启示是：在双语文本段落已经对齐的前提下，根据句末点号拆分段落中的每个句子；如果汉英双语文本中对应段落的句子数目相同，按照刚才发现的 1∶1对齐规律直接将序号相同的英汉句子对齐；如果句子数目不相同，则根据锚点（anchor）词等启发式信息，同时借助一部较好的双语词典对相对较多的句子进行归并对齐，详情

另文撰述。这种简单的基准(baseline)系统的句对齐准确率将不低于80%。

汉英对应语料库是建设翻译教学与研究平台，主要为人服务；与当前大多数对齐系统服务于统计机器翻译或实例机器翻译等自然语言处理工程项目不同。机器翻译等要求海量数据，以防止数据稀疏可能造成的统计错误；同时，一些遗漏或错误将随着数据量的扩大而被统计算法沉淀过滤。而句对对齐，质量要求更高，因此将简单有效的对齐和校对工具合二为一。

如图23.2所示，在界面最上方选择要处理的文件名；系统读取用户指定的中英文对译文本，根据基准处理策略对汉英语句子的粗对齐，并将粗对齐结果呈现在界面右上方“参考文档”的文本框中。用户可以在右下方的按钮区顺序或逆序对文本中每个句子进行遍历操作。当前正在编辑的句对中的中英文句子实例呈现在界面左部；例如本例中英句对比是5:3。现在假定用户已经将第1句汉语和第1句英文作为对齐链接；当前要将第2,3两句汉语句子和第2句英文句子做2:1的对齐，则可以通过选定汉语句子列表框(Listbox)中的第2～3行和英文句子列表框中的第2行，单击“对齐”按钮完成；执行结果将呈现在界面下部的“对齐句对”文本编辑框中。如果用户确认该结果，单击“保存”按钮，则将以新的对齐结果替代机器粗对齐结果；如果不认可还可以在该编辑框中重新处理，直到满意为止。针对图23.2中原始5:3的粗切分，经过人工干预后，将构成新的1:1,2:1,2:1的对齐格局。

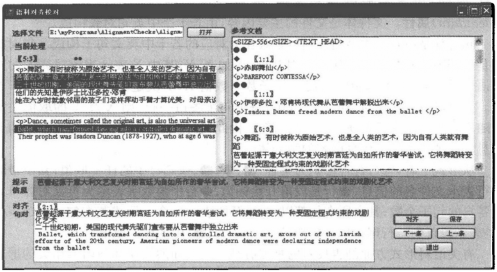


图23.2　语料库句子对齐校对界面

语料对齐校对辅助系统的好处在于：1)对齐校对界面无需用户输入过多的文字内容，例如句对中的中英文的句数由程序自动计算列表框中的用户选定行数确

定，对齐句子的文本复制也由系统自行完成，因此可以杜绝人工不熟练导致误操作形成的格式或字面上的错误；2)工作界面的核心区域集中在当前编辑的句对，并且一直在右侧提供包含当前句对的上下文语境，可以通过向前向后按钮实现遍历，这样防止由于文本章件过大，文字过多造成的遗漏检查；3)采用标准 Windows 鼠标操作，使得校对更快更方便，提高工作效率，加快工作进程。

23.4.2 中英双语语料前期人工对齐校对的再校对

语料库建设有一个不断深入的长期整理加工过程。在这个过程中，语料数量不断累积，处理手段也有更新。针对不同阶段创建加工的语料需要有一个一致化的处理。

双语对应语料库创建之初，曾经人工校对过一批经程序自动对齐的语料。但当时还没有开发合适的辅助校对工具，翻译专业人员在对句对对齐质量把关的同时，还是或多或少、不可避免地留下了因计算机操作不熟练或长期伏案工作、精力分散而导致的格式方面的错误。对这部分历史遗留语料的校对与上文介绍的一体化处理策略不同，着重在发现形式错误。

每一份英汉对齐语料都采用类 XML 表示设计，分为文件头(head)和文件体(body)等两个不同部分(常宝宝等 2003)。文件头包括该文件的标题、作者名、出版社、时间、语言及源语言和文体、领域等都柏林核心集中规定的元数据(*Metadata*)属性。文件体是文件的主要内容，由若干个中英句对簇构成。

句对簇是由句对类型引出随后紧跟互为对应的若干中英文句子构成。句对类型是由开闭符号“〖”“〗”括住的一对数值，形如 X:Y；其中 X 表示当前对齐的句对汉语句子个数，Y 表示相应的英文句子数。数值 X，Y 决定随后的中英文句子个数。按照对应语料库对齐的设计规范，每个句子各占一行，则 X 及 Y 的取值决定了随后若干行文本是否属于同一句对簇。因此句对簇是表示汉英对应句子数的元信息和汉英对英句子信息内容的结合体。

从形式上看，每行行首标识为集合{■、◆、●、<s>}中的任一取值；其中带■的行表示文件名的信息；●行是段落开始标识；◆行中的数字表示为英汉对齐单元中各自的句子个数；其余所有行都以<s>开始表示一个汉语或英文句子；<s><p>表示段首句。行尾标识为集合{】、</s>、〗、●}中的任一取值；

有关语料形式方面的再校对的重点有纯形式方面的错误和与内容相关的形式错误。前者包括 1)文本格式的规范化，譬如删除多余出来的空白行；2)将所有的软回车(\r)统一换作硬回车(\n)；3)将行尾标识与行首标识误拼接的重新拆分。后者包括由于对原始语料的对齐句对进行过调整，而元数据信息却没有修正，造成内容与形式的脱节，譬如原来语料中英文句对数为 1:2，在将第二个英语句子挪出

后，实质形成 1∶1的对应关系，但没有对句对应数更改造成元数据信息不能反映实际对齐句子的情况。

针对纯形式方面的错误，可以简单地采用全局统一替换的方式，如软硬回车的转变和空白行的删除，借助行首及行尾标识符也可以方便识别是否有行尾行首的误拼接；针对元数据信息和句对中英句子的校对采用如图 23.3 所示的处理方法解决。

输入：待校对文件；

输出：定位文件中句对数元数据信息与中英文实际对齐句子不匹配的内容

处理流程：

(1)打开待校对文件，指针指向文件头;设置计数单位CnNum，EnNum,初始为零;设置字符串CnSent,EnSent为空；

(2)当文件指针不指向文件尾，转(3)，否则转(6)：

(3)若当前行包含 [X:Y] ,为对应句数元数据信息， 其中X,Y为整数；

如果X!=CnNum或Y!=EnNum，则输出CnSent及EnSent；

置CnNum及EnNum为零，CnSent及EnSent为空串；转(2)；

(4)若当前行存在ASCII码大于128的字符，判断为汉语，CnNum累加；CnSent加入当前行；转(2)；

(5)若当前行不存在ASCII码大于128的字符，判断为英语，EnNum累加；EnSent加入当前行；转(2);

(6)关闭文件，退出应用程序。

图 23.3 人工校对语料元数据信息再校对处理描述

针对每一个待再校对的人工已校对文件，先处理纯形式方面的错误；再按照图 23.3 的操作流程，从前往后扫描文本，根据元数据信息的提示，获取随后句对簇中的对应句数，如果标识的句对数与实际句对数不一致，则将该部分内容输出，供工作人员检查校对。由此我们能得到一个既内容准确(因为经过早期的人工校对)，又形式合法(由于机器辅助的形式再校对)的新语料。

23.4.3 句子标注加工及校对

经过上述句对齐及校对之后，中英文句子之间的句对应关系可以确定，能够从任一语言的句子出发检索出相应的原文或译文。这一阶段的对应语料还是未经加工的生语料，例如汉语文本未经分词处理；而英文文本也没有经过词类赋码。因此，未经加工的生语料可能导致语料检索的准确度降低。譬如只能实现基于字而不是基于词的检索，如果要检索"华人"可能检索出"中华人民共和国"等这种包含带检索字串的内容。另一个不便之处是检索方式也受限，譬如无法要区分不同词

性的同一词形，像要分开检索作为介词的“把”和作为动词的“把”，如果语料未经分词和词类赋码，则根本就无能为力。因此有必要对已经句对齐的中英文句子进行适当的深加工，以满足用户不同颗粒度的检索需求。

在本测试语料中，我们利用中科院计算所的分词系统开源版本①，采用北京大学计算语言学研究所的汉语文本词性标注标记集，对汉语文本进行了词类赋码处理；与此类似，利用 Treetagger② 对英文文本进行了相应的词类赋码处理（Helmut 1994）。虽然这些赋码工具的准确性都达到了很高的程度，但我们设计的对应语料库是将被用户当作翻译教学与研究参考用的经典规范语料库，因此需要增加必要的校对工作。

我们设计了一个对句子词类赋码及元信息属性校对的辅助环境，工作界面如图 23.4 所示。在屏幕上方显著位置分别显示已赋码的中英文句子。下方显示该句所在篇章的篇名、作者名、篇章类型及写作时间的等元信息。其中，篇名及作者名不可编辑，只用于对其他可编辑元信息的检查提示。在文本框出现的中英文句子可以直接编辑修改，其他以多选框形式出现的元信息可以重新选择确定。由于句子标注的结果是“词＋码”的混合字符串形式，当句子较长时，词码混杂显示不太容易看出分词词类的结果，单击“不显示词类”按钮，则弹出只含有分词词形的句子结果，这样能够直观地观察分词结果。

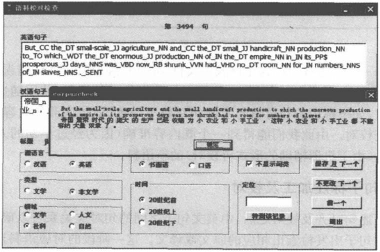


图 23.4　词类赋码及元信息加工校对

---

① http://www.nlp.org.cn/project/project.php?proj_id=6

② http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/

完成对中英文句子的词类赋码以及该句对的语种、源语言、文本类型及写作领域等元信息的校对之后，就可以考虑将汉英句对以及附着其上的元数据内容转化为关系型数据库，以利用数据库系统来进行组织管理。

23.4.4 数据库设计

为简化开发时间，我们在目前的汉英对应语料库系统中采用关系型数据库架构。这是因为数据库一来有较好的数据存储及管理及维护功能，二来也提供较好的检索服务机制，三来还具有方便的二次开发接口设计。在本项目中，双语对应语料库实际上设计为数据库中的一个表(table)。该表是由一个个汉英对应句子及其他属性构成的记录(record)组成。其中对每条记录而言，数据库定义为包括如表23.1所示的字段(field，属性)信息。

表23.1 数据库字段设计及说明

字段名	类 型	备 注
英文句	字符型(5 000)	一个英文句子不超过5 000字母
汉语句	字符型(2 000)	一个汉语句子不超过2 000汉字
源语言	字符型(1)	当前句子的源语言(英语或汉语)
类型	字符型(1)	当前句子所属的类型(取规定值之一)
文体	字符型(1)	当前句子所属的文体(取规定值之一)
语体	字符型(1)	当前句子所属的语体(取规定值之一)
时代	字符型(1)	当前句子所属的时代(取规定值之一)
段落ID	数值型	当前句子所在段落相对全文偏移量
句子ID	数值型	当前句子相对段落的偏移量
文件名	字符型(50)	当前句子所在文件名

在以上数据库字段的规划中，最主要的内容是“英文句”和“汉语句”，分别存入互有对应关系的完整的英语或汉语句子。为标准简单划一起见，当前对句子的判定以句末点号为准，如{。!?}等。英汉句子都分别进行了词类赋码，数据库记录中存放的是带码的词串，设计英文句子包含空格在内长度不超过5 000个字母，汉语句子最多不超过2 000个汉字。当前这只是个经验值，给可能的最长句子预留。如果万一遇到比预留长度更长的句子，则可以考虑将该句分拆以满足长度限制要求。

其他字段记录的是当前句对出处的文本属性及句子在原始文本中的偏移量等元信息。譬如可以从该句源语言的属性可以推知“英译汉”还是“汉译英”；从文体

属性可推知是文学体还是非文学体等。在对语料库进行检索或过滤时，可以根据这些字段属性加以各类综合条件的限制，从而得到所需结果。

23.5 小结

我们建造完成了迄今为止规模最大、经过专业人员精心校对的英汉双语对齐(句子层面)平行语料库。为使这一宝贵资源能够更好地服务于翻译研究和翻译教学，已经推出 Web 方式的原型测试检索系统，并得到了许多正面的建设性的意见和建议。

我们将继续搜集新的有代表性的双语语料，对其进行对齐、标注及校对等加工工序，并不断充实到双语对应语料库中，进一步扩大量的规模；同时根据用户相关反馈，修正语料库设计，进一步提升标注正确率和校对效率，使语料库的建设更贴近用户实际需求。

参 考 文 献

[1] Aarts, J. (1999), The description of language use [A], In H. Hasselgrd and S. Oksefjell (eds.), *Out of corpora: studies in honor of Stig Johansson* [C], Amsterdam: Rodopi, 3-20.

[2] Alcantar, C. (2007), Construction of a Special Corpus to Extract the Most Frequent Words in a Particular Field [J], *Essex Graduate Student Papers in Language and Linguistics* (9): 1-17.

[3] Anderman, G. and M. Rogers. (2003), *Translation Today: Trends and perspectives* [C], Clevedon: Multilingual Matters Ltd.

[4] Anderson, H. (1973), Abductive and deductive change [J], *Language* 49: 765-793.

[5] Bahumaid, S. (2006), Collocation in English-Arabic Translation [J], *Babel* 52(2): 133-152.

[6] Baker, M. (1992), *In Other Words: A Coursebook on Translation* [M], London: Routledge.

[7] Baker, M. (1993), Corpus Linguistics and Translation Studies: Implications and Applications [A], In M, Baker, G. Francis and E. Tognini-Bonelli (eds.), *Text and Technology: In Honour of John Sinclair* [C], Philadelphia and Amsterdam: John Benjamins, 233-250.

[8] Baker, M. (1995), Corpora in translation studies: An overview and some suggestions for future research [J], *Target* 7(2): 223-243.

[9] Baker, M. (1996), Corpus-based translation studies: the challenges that lie ahead [A], In H. Somers (ed.), *Terminology, LSP, and translation* [C], Amsterdam: John Benjamins, 175-186.

[10] Baker, M. (1998a), Réexplorer la langue de la traduction: une approche par corpus [J], *Meta* 43(4): 480-485.

[11] Baker, M. (1998b), *Routledge Encylopedia of Translation Studies* [Z], London and New York: Routledge.

[12] Baker, M. (1999), The role of corpora in investigating the linguistic behaviour of professional translation [J], *International Journal of Corpus Linguistics* (2): 281-298.

[13] Baker, M. (2000), Towards a Methodology for Investigating the Style of a Literary Translator [J], *Target* 12(2): 241-266.

[14] Baker. M. (2004), The treatment of variation in corpus-based translation studies [A], In K. Aijmer and H. Hasselgrd (ed.), *Translation and Corpora* [C], Acta Universitatis Gothoburgensis, 7-17.

[15] Barlow, M. (2000), *Parallel Text in Language Teaching* [C], In S. P. Botley, A. M. McEnery and A. Wilson (eds.), *Multilingual Corpora in Teaching and Research* [C], Amsterdam: Rodopi, 106-115.

[16] Barnbrook, G., P. Danielsson and M. Mahlberg (eds.), (2006), *Meaningful Texts: The Extraction of Semantic Information from Monolingual and Multilingual Corpora* [C], London and New York: Continnum.

[17] Bell, R. T. (1991), *Translation and Translating: Theory and Practice* [M], London and New York: Longman.

[18] Benson, P. (2001), *Teaching and Researching Autonomy in Language Learning* [M], Harlow: Pearson Education.

[19] Berger, A., Brown, P. and S. Peitra, et al. (1994), The Candide System for Machine Translation [A], In *Proceedings of the Workshop on Human Language Technology* [C], Plainboro, NJ. 1994, 157-162.

[20] Bernardini, S. (1997), Trainee translator's perspective on corpora [A], *Paper presented at Corpus Use and Learning to Translate* held at Bertinoro, November, Available on-line at URL http://www.sslmit.unibo.it/introduz.htm.

[21] Bernardini, S. (2004a), Corpora in the classroom: An overview and some reflections on future developments [A], In J. M. Sinclair (ed.), *How to Use Corpora in Language Teaching* [C], Philadelphia: John Benjamins, 15-36.

[22] Bernardinin, S. (2004b), Corpus-aided language pedagogy for translator education [A], In K. Malmjær (ed.), *Translation in Undergraduate Degree Programmes* [M], Amsterdam: John Benjamins, 97-112.

[23] Bernardini, S. (2005), Reviving old ideas: parallel and comparable analysis in translation studies—with an example from translation stylistics [A], In K. Aijmer and C. Alvstad (eds.), *New Tendencies in Translation Studies. Selected Papers from a Workshop Goteborg 12 December 2003* [C], Guteborg: Acta Universitatis Gothoburgensis, 5-18.

[24] Bernardini, S. and F. Zanettin (2004), When is a universal not a universal? Some limits of current corpus-based methodologies for the investigation of translation universals [A], In A. Mauranen and P. Kujamaki (eds.), *Translation Universals—Do They Exist?* [C], Amsterdam: John Benjamins, 51-62.

[25] Bernardini, S., et al. (2003), An introduction. In F. Zanettin et al. (eds.) *Corpora in Translator Education* [C], Manchester: St. Jerome Publishing.

[26] Blum-Kulka, S. and E. A. Levenston. (1983), Universals of lexical simplification [A], In C. Furch and G. Kasper (eds.), *Strategies in Interlanguage Communication* [C], London and New York: Longman, 119-139.

[27] Blum-Kulka, S. (1986), Shifts of cohesion and coherence in translation [C], In J. House and S. Blum-Kulka (eds.), *Interlingual and Intercultural Communication. Discourse and Cognition in Translation and Second Acquisition Studies* [C], Tübingen: Gunter Narr, 61-71.

[28] Boase-Beier, J. (2003), Mind Style Translated [J], *Style* 37(3): 253-265.

[29] Boase-Beier J. (2004), Translation and style: A brief introduction [J], *Language and Literature* 13 (3): 9-11.

[30] Boase-Beier, J. (2006), *Stylistic Approaches to Translation* [M], Manchester: St. Jerome.

[31] Bosseaux, C. (2004), Point of View in Translation: A corpus-based study of French translations of Virginia Woolf's To the Lighthouse [J], *Across Languages and Cultures* 5 (1): 107-122.

[32] Bowker, L. (1998), Using specialized monolingual native-language corpora as a translation resource: A pilot study [J], *Meta* (43)4: 631-651.

[33] Bowker, L. (2003), Corpus-based applications for translator training: Exploring the possibilities [A], In S. Granger et al. (eds.), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies* [C], Amsterdam: Rodopi, 169-183.

[34] Brinton, L. J. and E. C. Traugott (2005), *Lexicalization and Language Change* [M], Cambridge: Cambridge University Press.

[35] Brown, P., V. Pietra and S. Pietra (1993), The Mathematics of Statistical Machine Translation: Parameter Estimation [J], *Computational Linguisitcs* (19): 263-311.

[36] Catford, J. C. (1965), *A linguistic theory of translation* [M], Oxford: OUP.

[37] Chanell, J. (1980), Applying Semantic Theory to Vocabulary Teaching [J], *ELT Journal*, 35(2), 115-122.

[38] Li, C. and S. Thompson (1981), *Mandarin Chinese: A Functional Grammar* [M], Berkley: University of California Press.

[39] Chen, H. (1994), The contextual analysis of Chinese sentences with punctuation marks [J], *Literary and Linguistic Computing* 9(4): 281-289.

[40] Chesterman, A. (1997), *The memes of translation* [M], Amsterdam: John Benjamins.

[41] Chesterman, A. (2000), A Causal model for translation studies [A], In M. Olohan (ed.), *Intercultural faultlines. Research models in translation studies I: Textual and cognitive aspects* [C], Manchester: St Jerome, 15-27.

[42] Chesterman, A. (2004a), Beyond the particular [A], In A. Mauranen and P. Kujamaki (eds.), *Translation Universals: Do They Exist?* [C], Amsterdam: John Benjamins, 33-

49.

[43] Chesterman, A. (2004b), Hypothesis about translation universals [A], In G. Hansen, K. Malmkjaer and D. Gile (eds.), *Claims, Changes and Challenges in Translation Studies* [C], Amsterdam: John Benjamins, 1-13.

[44] Chesterman, A. (2005), Toward Consilience? [A], In K. Aijmer and C. Alvstad (eds.), New Tendencies in Translation Studies [M], Göteborg: Acta Universitatis Gothoburgensis, 19-27.

[45] Chujo, K., M. Utiyama and K. Oghigian. (2006), Selecting Level-Specific Kyoto Tourism Vocabulary Using Statistical Measures [R], *Paper presented at Conference of 2006 International Teaching and Learning*: New Aspect of English Language Learning in Taiwan. Available online at URL: http://www2.nict.go.jp/x/x161/members/mutiyama/pdf/Kyoto-Article-Chujo-Utiyama-Oghigian.pdf.

[46] Cowie, P. (1994), Phraselogy. In R. Asher (ed.), *The Encyclopedia of Language and Linguistics* [C], Oxford: Pergamon, 3168-3171.

[47] Craciunescu, O., et al. (2004), Machine Translation and Computer-Assisted Translation: a New Way of Translating? [J], *Translation Journal* 8(3) http://translationjournal.net/journal//29computers.htm

[48] Crisafulli, E. (2002), The quest for an eclectic methodology of translation description [A], In T. Hermans (ed.), *Crosscultural Transgressions: Research Models in Translation Studies* II [C], Manchester: St. Jerome, 26-43.

[49] Croft, W. (1990), *Typology and Universals* [M], Cambridge: CUP.

[50] Crystal, D. (1999), *The Penguin Dictionary of Language* [Z], London: Penguin Books.

[51] Davidse, K. and L. Heyvaert. (2003), On the middle construction in English and Dutch. In Granger, S., J. Lerot and S. Petch-Tyson. (eds.), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam: Rodopi, 57-73.

[52] Deutscher, G. (2002), On the misuse of the notion of 'abduction' in linguistics [J], *Journal of Linguistics* 38: 469-485.

[53] Dickinson, L. (1987), *Self-instruction in Language Learning* [M], Cambridge: Cambridge University Press.

[54] Duff. Alan. (1981), *The Third Language: Recurrent problems of translation into English* [M], Oxford: Pergamon Press.

[55] Ebeling, J. (1998), Contrastive linguistics, translation, and parallel corpora [J], *Meta* 43 (4): 602-615.

[56] Englund-Dimitrova, B. (2003), Explicitation in Russian-Swedish translation: sociolinguistic and pragmatic aspects [A], In B. Englund-Dimitrova and A. Pereswetoff-Morath (eds.), *Swedish Contributions to the Thirteenth International Congress of Slavists*, Ljubljana,

15-21 August, 2003 [C], Lund: Lund University, 21-31.

[57] Englund-Dimitrova, B. (2005). *Expertise and Explicitation in the Translation Process* [M], Amsterdam: John Benjamins.

[58] Eskola, S. (2004). Untypical frequencies in translated language: A corpus-based study on a literary corpus of translated and non-translated Finnish [A], In A. Mauranen and P. Kujamæki (eds.), *Universals in Translation. Do they exist?* [C], Amsterdam: John Benjamins, 83-99.

[59] Firth, J. (1957). *Papers in Linguistics* [M], London: OUP.

[60] Francis, W. N. and H. Kuaera (1982), *Frequency Analysis of English Usage: Lexicon and Grammar* [M], Boston: Houghton Mifflin.

[61] Frankenberg-Gacia, A. (2004), Are translations longer than source texts? A corpus-based study of explicitation [A], *Paper presented at the Third International CULT* (Corpus Use and Learning to Translate) Conference, Barcelona, 22-24 January, 2004, http://www.linguateca.pt/Repositorio/Frankenberg-Garcia2004.doc (accessed 07/04/2006)

[62] Frawley, W. (1984), Prolegomenon to a theory of translation [A], In W, Frawley (ed.), *Translation: Literary, linguistic and philosophical perspectives* [C], London and Toronto: Associated University Press, 159-175.

[63] Friedbichler, I. and M. Friedbichler, (1997), The potential of domain-specific target-language corpora for the translator's workbench [A], Paper presented at the Conference of Corpus use and learning to translate held at Bertinoro, Novembe 1997.

[64] Gaussier eric, et al. (2000), Term Alignment in Use: Machine-Aided Human Translation [A], In Véronis, Jean (ed.), *Parallel Text Processing: Alignment and Use of Translation Corpus* [C], Cordrecht: Kluwer Academic Publishers, 253-274.

[65] Gellerstam, M. (1986), Translationese in Swedish novels translated from English [A], In L. Wollin and H. Lindquist (eds.), *Translation Studies in Scandinavia* [C], Lund: CWK Gleerup: 88-95.

[66] Granger, S., J. Lerot and S. Petch-Tyson. (2003), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies* [M], Amsterdam: Rodopi.

[67] Halverson, Sandra. (2003), The cognitive basis of translation universals [J], *Target* 15 (2): 197-241.

[68] Hansen, S. and Teich, E. (2001), Multi-layer analysis of translation corpora: methodological issues and practical implications [A], In D. Cristea, N. Ide, D. Marcu and M. Poesio (eds), *Proceedings of EUROLAN 2001 Workshop on Multi-layer Corpus-based Analysis* [C], Iasi, Romania, 44-55.

[69] Hantson, A. (2003). English gerund clauses and Norwegian *det* + infinitive/at clause constructions. In Granger, S., J. Lerot and S. Petch-Tyson. (eds.), *Corpus-based*

Approaches to Contrastive Linguistics and Translation Studies. Amsterdam: Rodopi, 75-90.

[70] Hatim, B. and I. Mason. (1990/2001), *Discourse and the Translator* [M], Shanghai: Shanghai Foreign Language Education Press.

[71] Hatim, Basil. (2001), *Teaching and Researching Translation* [M], Harlow: Pearson Education.

[72] Helmut Schmid, (1994), Probabilistic Part-of-Speech Tagging Using Decision Trees [A], In *International Conference on New Methods in Language Processing*. [C], Manchester, UK: 44-49

[73] Hermans, T. (ed.) (2002), *Crosscultural Transgressions: Research models in translation studies* Ⅱ: *Historical and ideological issues* [C], Manchester: St. Jerome.

[74] Hermans, T. (1996), The Translator's Voice in Translated Narrative [J], *Target* 8(1): 23-48.

[75] Hermans, T. (1999), *Translation in systems* [M], Manchester: St Jerome.

[76] Hermans, T. (2002), Paradoxes and aporias in translation and translation studies [A], In A. Riccardi (ed.), *Translation Studies: Perspectives on an Emerging Discipline* [C], Cambridge: Cambridge University Press: 10-23.

[77] Hobbs, J. (2004), Abduction in natural language understanding [A], In L. R. Horn and G. Ward (eds.), *The handbook of pragmatics* [C], London and New York: Blackwell, 724-741.

[78] Holmes, J. S. (1972/1988). The name and nature of translation studies [A], In J. S. Holmes. *Translated! Papers on literary translation and translation studies* [M], Amsterdam: Rodopi, 67-80.

[79] Hoover, D. L. (2007), Corpus Stylistics, Stylometry, and the Styles of Henry James [J], *Style* 41(2): 174-203.

[80] Hopkinson, C. (2007), Factors in linguistic interference: A case of study in translation [J], *Skase Journal of Translation and Interpretation* 2(1): 13-23.

[81] Hudson, R. A. (1996/2000), *Sociolinguistics* [M], Beijing: Foreign Language Teaching and Research Press.

[82] Huiming Yang. (2006), Blended learning—a new business English solution [J], *US-China Foreign Language*, (9):67-71.

[83] Hunston, S. and G. Francis. (2000), *Pattern grammar* [M], Amsterdam: John Benjamins.

[84] Hunston, S. (2002), *Corpora in Applied Linguistics* [M], Cambridge: Cambridge University Press.

[85] Johansson, S. and K. Hofland. (1994), Towards an English-Norwegian Parallel Corpus [A], In U. Fries, G. Tottie and P. Schneider (eds), *Creating and Using English Language Corpora* [C], Amsterdam: Rodopi, 25-37.

[86] Johansson, S. (1991), Times Change and so do Corpora [A], In K. Aijmer and B. Altenberg (eds), *English Corpus Linguistics: Studies in Honour of Jan Svartvik* [C], London and New York: Longman, 305-314.

[87] Johansson, S. (1998), On the role of corpora in cross-linguistic research [A], In S. Johansson and S. Oksefjell (eds), *Corpora and Cross-Linguistic Research* [C], Amsterdam: Rodopi, 1-24.

[88] Johansson, S. (2003), Reflections on corpora and their uses in cross-linguistic research [A], In F. Zanettin, S. Bernardini and D. Stewart (eds), *Corpora in Translator Education* [C], Manchester: St. Jerome, 135-144.

[89] Johansson, S. (2004), Why change subject? On changes in subject selection in translation from English into Norwegian [J], *Target* 16 (1): 29-52.

[90] Johansson, S. (2007), *Seeing Through Multilingual Corpora* [M], Amsterdam: John Benjamins.

[91] Johns, T. and King P. (1991), *Classroom Concordancing* [C], ELR Journal, New Series 4.

[92] Kennedy, G. (1998), *An Introduction to corpus linguistics*. London: Longman.

[93] Kenny, D. (1998), Creatures of habit? What translators usually do with words [J], *Meta* 43 (4): 1-9.

[94] Kenny, D. (1999), Translators at play: exploitations of collocational norms in German-English translation [A], In B. Dodd (ed.), *Working with German Corpora* [C], Birmingham: University of Birmingham Press, 143-160.

[95] Kenny, D. (2000), Lexical hide-and-seek: looking for creativity in a parallel corpus [A], In M. Olohan (ed.), *Intercultural faultlines. Research models in translation studies I: Textual and cognitive aspects* [C], Manchester: St Jerome, 93-104.

[96] Kenny, D. (2001), *Lexis and Creativity in Translation: A Corpus-based Study*. [M], Manchester: St. Jerome Publishing.

[97] Kenny, D. (2005), Parallel corpora and translation studies: old questions, new perspectives? Reporting that in Gepcolt: a case study [A], In G. Barnbrook, P. Danielsson and M. Mahlberg (eds.), *Meaningful Texts: The Extraction of Semantic Information from Monolingual and Multilingual corpora* [C], London: Continuum, 154-165.

[98] Klaudy, K. and K. Károly (2005), Implicitation in Translation: Empirical Evidence for Operational Asymmetry in Translation [J], *Across Languages and Cultures*, 6 (1): 13-28.

[99] Klaudy, K. (1998/2004), Explicitation [A], In M. Baker (ed.), *Routledge Encyclopedia of Translation Studies* [Z], Shanghai: Shanghai Foreign Language Education Press, 80-84.

[100] Koehn, P. (2005), *Europarl: A Parallel Corpus for Statistical Machine Translation*

[A], Proceedings of 10th Machine Translation Summit [C].

[101] Kübler, N. (2003), Corpora and LSP Translation [A], In F. Zanettin, S. Bernardini and D. Stewart (eds), *Corpora in Translator Education* [C], Manchester: St. Jerome, 25-42.

[102] Kuhn, T. (1962), *The structure of scientific revolutions* [M], Chicago: University of Chicago Press.

[103] Laham, D. (1997), Latent Semantic Analysis Approaches to Categorization [A], In M. G. Shafto and P. Langley (eds.), *Proceedings of the 19th Annual Conference of the Cognitive Science Society*. [C] Hillsdale, NJ: Lawrence Erlbaum Associates, 979.

[104] Landauer, T. K., P. W. Foltz, and D. Laham (1998), Introduction to Latent Semantic Analysis [J], *Discourse Processes* (25): 259-284.

[105] Laviosa, S. (1997), How comparable can comparable corpora be? [J], *Target* 9: 289-319.

[106] Laviosa, S. (1998a), The Corpus-based Approach: A New Paradigm in Translation Studies [J], *Meta* (43/4): 474-479.

[107] Laviosa, S. (1998b), The English comparable corpus: a resource and a methodology [A], In L. Bowker, M. Cronin, D. Kenny and J. Pearson (eds.), *Unity in Diversity? Current Trends in Translation Studies* [C], Manchester: St. Jerome: 101-112.

[108] Laviosa, S. (1998c), Core patterns of lexical use in a comparable corpus of English narrative prose [J], *Meta* 43(4): 557-570.

[109] Laviosa, S. (2002), *Corpus-based Translation Studies: Theory, Findings and Applications* [M], Amsterdam: Rodopi.

[110] Laviosa, S. (2003), Corpora and translation studies [A], In S. Granger, Jacques Lerot and Stephanie Petch-Tyson (eds.), *Coprus-based approaches to contrastive linguistics and translation studies* [C], Amsterdam: Rodopi: 45-54.

[111] Laviosa-Braithwaite, S. (1998/2004), Universals of translation [A], In M. Baker (ed.), *Routledge Encyclopedia of Translation Studies* [Z], Shanghai: Shanghai Foreign Language Education Press, 288-291.

[112] Leech, G. and M. Short (1981/2001), *Style in Fiction: A Linguistic Introduction to English Fictional Prose* [M], Beijing: Foreign Language Teaching and Research Press.

[113] Leech, G., P. Rayson and A. Wilson. (2001), *Word Frequencies in Written and Spoken English based on the British National Corpus* [M], Harlow andLondon: Pearson Education.

[114] Leech, G. (1997), Teaching in language corpora: a convergence [A], In G Knowles, T. McEnery, S. Fligelstone, and A. Wichman (eds.), *Teaching and language Corpora* [C], London: Longman, 1-22.

[115] Li, D. (2004), Trustworthiness of Think-aloud Protocols in the Study of Translation

Process [J], *International Journal of Applied Linguistics* (3): 301-313.

[116] Maia, B. (1998), Word order and the first person singular in Portuguese and English [J], *Meta* 43(4): 589-601.

[117] Maia, B. (2003), Some languages are more equal than others: training translators in terminology and information retrieval using comparable and parallel corpora [A], In F. Zanettin, S. Bernardini and D. Stewart. *Corpora in Translator Education* [C], Manchester: St. Jerome, 43-53.

[118] Malmkjaer, K. (2003), What happened to God and the angels: An exercise in translational stylistics [J], *Target* 15(1): 37-58.

[119] Malmkjaer, K. (2004), Translational stylistics: Dulcken's translations of Hans Christian Andersen [J], *Language and Literature* 13(1): 13-24.

[120] Malmkjær, K. (2005), *Linguistics and the Language of Translation* [M], Edinburgh: EUP

[121] Manning, C. D. and H. Schütze (2005),《统计自然语言处理基础》[M],苑春法,等译.北京:电子工业出版社.

[122] Marco, J. (2004), Translating style and styles of translating: Henry James and Edgar Allan Poe in Catalan [J], *Language and Literature* 13(1): 73-90.

[123] Mason, Ian. (2001), Translator behavior and language usage: Some constraints on contrastive studies [J], *Hermes* (26): 65-80.

[124] Mauranen, A. and P. Kujamæki (eds.), (2004), *Translation Universals: Do They Exist* [C], Amsterdam: John Benjamins.

[125] Mauranen, A. (2002), Where's cultural adaptation. A corpus-based study on translation strategies [A], http://www.intralinea.it/(accessed 27/11/2005)

[126] Monzó, E. (2003), Corpus-based teaching: The use of original and translated texts in the training of legal translators [J], *Translation Journal*, (7)4. Available on-line at http://accurapid.com/journal/26edu.htm.

[127] McEnery, T. and M. Oakes. (1996), Sentence and word alighment in the CRATER project [A], In J. Thomas and M. Short (eds.) *Using Corpora for Language Research* [C], London: Longman, 211-231.

[128] McEnery, T. and R. Xiao. *The Lancaster Corpus of Mandarin Chinese* (*LCMC*) [O] http://www.lancs.ac.uk/fass/projects/corpus/LCMC/. Last Retrieved on September 15, 2009.

[129] Millán-Varela, C. (2004), Hearing voices: James Joyce, Narrative Voice and Minority Translation [J], *Language and Literature* 13(1): 37-54.

[130] Monzó, E. (2003), Corpus-based teaching: The use of original and translated texts in the training of legal translators [J], *Translation Journal* (4), Available on-line at http://accurapid.com/journal/26edu.htm.

[131] Nagao, M. (1984), A framework of a mechanical translation between Japanese and English by analogy principle [A], In A. Elithorn and R. Banerji (eds.) *Artificial and Human Intelligence* [C], Amsterdam: North-Holland, 173-180.

[132] Newman, A. (1980), *Mapping translation equivalence* [M], Leuven: ACCO.

[133] Nida, E. A. and C. Taber. (1969), *The Theory and Practice of Translation* [M], Leiden: E. J. Brill.

[134] Nida, E. A. (1964/2004), *Toward a Science of Translating* [M], Shanghai: Shanghai Foreign Language Teaching Press.

[135] Nord, C. (2005), *Text Analysis in Translation: Theory, methodology, and didactic application of a model for translation-oriented text analysis* (2nd edition) [M], Amsterdam: Rodopi.

[136] Olohan, M. and M. Baker (2000), Reporting that in Translated English: Evidence for Subconscious Processes of Explicitation [J], *Across Languages and Cultures* 1(2): 141-158.

[137] Olohan, M. (2001), Spelling out the Optionals in Translation: A corpus study [J], *UCREL Technical Papers* 13: 423-432.

[138] Olohan, M. (2002), Leave it out! Using a comparable corpus to investigate aspects of explicitation in translation [J], Cadernos de Tradução 9: 153-169.

[139] Olohan, M. (2003), How frequent are the contractions? A study of contracted forms in the Translational English Corpus [J], *Target* 15 (1), 59-89.

[140] Olohan, M. (2004), *Introducing Corpora in Translation Studies* [M], London and New York: Routledge.

[141] Ǿverås, L. (1998), In Search of the Third Code: An Investigation of Norms in Literary Translation [J], *Meta*, 43 (2): 571-588.

[142] Paloposki, O. (2001), Enriching translations, simplified language? An alternative viewpoint to lexical simplification [J], *Target* 13 (2): 265-288.

[143] Papineni, K., and S. Roukos (2002), Bleu: a Method for Automatic Evaluation of Machine Translation. [C] *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia, July 2002, 311-318.

[144] Pearson, J. (2003), Using parallel texts in the translator training environment [A], In F, Zanettin, S, Bernardini and D, Stewart (eds.) *Corpora in Translator Education* [C], Manchester: St. Jerome, 15-24.

[145] Peirce, C. S. (1931-1966), *The collected papers of Charles Sanders Peirce* [1839-1914] [C], (8 vols.), Cambridge, MA: Harvard University Press.

[146] Popper, Karl. (1979), *Objective Knowledge: An Evolutionary Approach* [M] (revised edition). Oxford: Oxford University Press.

[147] Puurtinen, T. (1998), Syntax, readability and ideology in children's literature [J],

Meta 43 (4): 524-533.

[148] Puurtinen, T. (2003a), Genre-specific features of translationese? Linguistic differences between translated and non-translated Finnish children's literature [J], *Literary and Linguistic Computing*. 18 (4): 389-406.

[149] Puurtinen, T. (2003b), Non-finite constructions in Finnish children's literature: features of translationese contradicting translation universals [A], In S. Granger, J. Lerot and S. Petch-Tyson (eds.), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies* [C], Amsterdam: Rodopi, 141-154.

[150] Puurtinen, T. (2004), Explicitation of clausal relations: A corpus-based analysis of clause connectives in translated and non-translated Finnish children literature [A], In Anna Mauranen and Pekka Kujam Ki (eds.), (2004), *Translation Universals: Do They Exist?* [C] Amsterdam: John Benjamins, 165-176.

[151] Pym, A. (2003), Redefining Translation Competence in an Electronic Age: In Defense of a Minimalist Approach [J], *Meta* 43(4): 481-497.

[152] Quirk, R., S. Greenbaum, G. Leech and J. Svartvik (eds.) (1972), *A Grammar of Contemporary English* [M], London: Longman.

[153] Riccardi, A. (2002), *Translation Studies: Perspectives on an emerging discipline* [M], Cambridge: Cambridge University Press.

[154] Santos, D. and A. Frankenberg-Garcia. (2007), The corpus, its users and their needs, a user-oriented evaluation of COMPARA [J], *International Journal of Corpus Linguistics* 12(3): 335-374.

[155] Santos, D. (1995), *On grammatical translationese* [A], In K. Koskenniemi (ed.), Short Papers Presented at the Tenth Scandinavian Conference on Computational Linguistics [C], Helsinki: 59-66.

[156] Schmied, J. and H. Schäffler. (1997), Explicitness as a universal feature of translation [A], In M. Ljung (ed.), *Corpus-based Studies in English. Papers from the 17th International Conference on English Language Research on Computerized Corpora (ICAME 17), Stockolm, May 15-19, 1996* [C], Amsterdam: Rodopi.

[157] Semino, E. and M. Short. (2004), Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing [M], London: Routledge.

[158] Serpellet, N. (2002). Mandative constructions in English and their equivalents in French: applying a bilingual approach to the theory and practice of translation [A], In B. Kettemann and G. Marko (eds.), *Teaching and Learning by Doing Corpus Analysis. Proceeding of the Fourth International Conference on Teaching and Language Corpora* [C], Amsterdam, 345-360.

[159] Shlesinger, M. (1998). Corpus-based interpreting studies as an offshoot of corpus-based translation studies [J], *Meta* (43)4: 486-493.

[160] Shuttleworth, M. and M. Cowie (1997), *Dictionary of Translation Studies* [Z], Manchester: St. Jerome Publishing.

[161] Sinclair, J. M. (1991/1999), *Corpus, concordance, collocation* [M], Shanghai: Shanghai Foreign Language Education Press.

[162] Sinclair, J. M. (1996), *How to Use Corpora in Language Teaching* [M], Amsterdam: John Benjamins.

[163] Snell-Hornby, M. (2006), The Turns of Translation Studies: New paradigms or shifting viewpoints [M], Amsterdam: John Benjamins.

[164] Stubbs, M. (1986), Language development, lexical competence and nuclear vocabulary [A], In K. Durkin (ed) *Language Development in the School Years* [C], Croom Helm.

[165] Stubbs, M. (1993), British tradition in text analysis from Firth to Sinclair [A], In M. Baker, G. Francis and E. Tognini-Bonelli (eds), *Text and Technology: In Honour of John Sinclair* [C], Amsterdam: John Benjamins, 1-33.

[166] Tirkkonen-Condit, S. (2002), Translationese—A myth or an empirical fact? A study into the linguistic identifiablity of translated language [J], *Target* 14 (2), 207-220.

[167] Tirkkonen-Condit, S. (2004), Unique items—over- or under-represented in translated language [A], In A. Mauranen and P. Kujamæki (eds.), *Translation Universals—Do They Exist?* [C]Amsterdam: John Benjamins, 177-184.

[168] Tognini-Bonelli, E. (2001), *Corpus Linguistics at Work* [M], Amsterdam: John Benjamins.

[169] Toury, G. (1978), The Nature and Role of Norms in Literary Translation [A], In J. S. Holmes, J. Lambert and R. van den Broeck, (eds.), *Literature and Translation. New Perspectives in Literary Studies* [C], Leuven: Acco, 83-100.

[170] Toury, G. (1980), *In Search of a Theory of Translation* [M], Tel Aviv: The Porter Institute for Poetics and Semiotics.

[171] Toury, G. (1995/2001), Descriptive Translation Studies and Beyond [M], Shanghai: Shanghai Foreign Language Education Press.

[172] Toury, G. (2004), Probabilistic explanations in translation studies: universals—or a challenge to the very concept [A], In G. Hansen, K. Malmkjaer and D. Gile (eds.), *Claims, Changes and Challenges in Translation Studies. Selected Contributions for the EST Congress, Copenhagen, 2001* [C], Amsterdam: John Benjamins, 15-25.

[173] Trosborg, A. (ed.), (1997), *Text Typology and Translation* [C], Amsterdam: John Benjamins,

[174] Trudgill, P. (1983), Sociolinguistics: An Introduction to Language and Society [M], Harmondsworth: Penguin Books.

[175] Tsao, F. (1979), *A Functional Study of Topic in Chinese* [M], Taipei: Student Book Co.

[176] Turney, P. D. (2005), Measuring semantic similarity by latent relational analysis. [C] *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence* (IJCAI-05), Edinburgh, Scotland, 1136-1141.

[177] Tymoczko, M. (1998), Computerized corpora and the future of translation studies [J], *Meta* 43(4): 652-660.

[178] Tymoczko, M. (2005), Enlarging translation theory: Integrating Non-western thoughts about translation [A], In T. Hermans. *Translating Others* [C], Manchester: St. Jerome, 13-32.

[179] Tytler, A. (1907/2007), *Essays on Principles of Translation* [M], Beijing: Foreign Language Teaching and Research Press.

[180] Ure, J. (1971), Lexical density and register differentiation [A], In G. E. Perren and J. L. M. Trim (eds.), *Applications of linguistics: selected papers of the Second International Congress of Applied Linguistics* [C], Cambridge: Cambridge University Press, 443-452.

[181] Vanderauwera, R. (1985), Dutch Novels Translated Into English. Amsterdam: Rodopi.

[182] Venuti, L. (1995), The translator's Invisibility: A History of translation [M], London: Routledge.

[183] Véronis, J. (ed.), (2000), *Parallel Text Processing: Alignment and Use of Translation Corpus* [C], Cordrecht: Kluwer Academic Publishers.

[184] Veronis, J. (2000), From the Rosetta stone to the information society: a survey of parallel text processing [A], In J. Véronis (ed.), *Parallel Text Processing: Alignment and Use of Translation Corpus* [C], Cordrecht: Kluwer Academic Publishers, 1-24.

[185] Vinay, J. and J. Darbelnet. (1958/1995), *Comparative Stylistics of French and English. A Methodology for Translation* [M], Translated and edited by J. C. Sager and M. -J. Hamel. Amsterdam: John Benjamins.

[186] Wales, K. (2001), *A Dictionary of Stylistics* [Z], London: Longman.

[187] Wichmann, A., S. Fligelstone, A. McEnery, and G. Knowles (eds.), (1997), *Teaching and Language Corpora* [C], London: Longman.

[188] Wilkinson, M. (2005), Discovering Translation Equivalents in a Tourism Corpus by Means of Fuzzy Searching [J], *Translation Journal* (9): 1-10.

[189] Wilkinson, M. Compiling a specialized corpus to be used as a translation aid [OL], available online at URL http://www.joensuu.fi/hallinto/jopke/dokumentit/Wilkinson.doc

[190] Wynne, M. (2005), Stylistics: corpus approaches [OL], Available at: www.pala.ac.uk/resources/sigs/corpus-style/Corpora_stylistics.pdf (accessed 22-10-2008)

[191] Xiao, Z. and T. McEnery (2002), A corpus-based approach to tense and aspect in English-Chinese translation [R], *Paper presented at International Symposium on*

Contrastive and Translation Studies between Chinese and English. Shanghai, 2002.

[192] Zanettin, F. (1998), Bilingual comparable corpora and the training of translators [A], *Meta* (4): 4-14.

[193] Zanettin, F., S. Bernardini and D. Stewart (eds), (2003), *Corpora in Translator Education* [C], Manchester: St. Jerome.

[194] 常宝宝、柏晓静.(2003),北京大学汉英双语语料库标记规范[J],《汉语语言与计算学报》13(2):195-214.

[195] 陈定安.(1995),《英汉比较与翻译》[M],北京:中国对外翻译出版公司.

[196] 陈国华.(1997),论莎剧重译(下)[J],《外语教学与研究》(3):48-54.

[197] 陈宏薇.(2004),《新编汉英翻译教程》[M],上海:上海外语教育出版社.

[198] 陈瑞清.(2004),汉译文本的形合趋势:以语料库为本的翻译学研究[A],《第九届口笔译教学研讨会论文集》[C],台南:长荣大学翻译学系.

[199] 程雨民(编著),(1989),《英语语体学》[M],上海:上海外语教育出版社.

[200] 董红源.(2002),汉语代词"他"的句内照应规则研究[MA],北京大学汉语系硕士论文.

[201] 范仲英.(1997),《实用翻译教程》[M],北京:外语教学与研究出版社.

[202] 封一函.(2001),教室网络申的交互式翻译教学[J],《中国翻译》(2):37-39.

[203] 封宗颖、邵志洪.(2004),英汉第三人称代词深层回指对比与翻译[J],《外语学刊》.(5):95-100.

[204] 冯庆华.(1997),《实用翻译教程》[M],上海:上海外语教育出版社.

[205] 冯志伟.(1996),《自然语言的计算机处理》[M],上海:上海外语教育出版社.

[206] 冯志伟.(2001),中国语料库研究的历史与现状[A],《国际汉语计算机会议ICCC2001论文集》[C],新加坡.

[207] 冯志伟.(2007),基于经验主义的语料库研究[J],术语标准化与信息技术》(1):29-36.

[208] 高照明.(2004),汉英近义句翻译检索,载王克非等著,《双语对应语料库:研制与应用》[M],北京:外语教学与研究出版社,69-79.

[209] 郭红.(2004),计算机辅助翻译教学的一种尝试[J],《外语界》(5):54-61.

[210] 何兆武、李约瑟(译).(1963/1980),《西方哲学史》(罗素原著)[M],北京:商务印书馆.

[211] 贺显斌.(2003),英汉翻译过程中的显化现象[J],《解放军外国语学院学报》(4):63-66.

[212] 贺阳.(2004),从现代汉语介词中的欧化现象看间接语言接触[J],《语言文字应用》(4):82-89.

[213] 胡开宝、朱一凡.(2008),基于语料库的莎剧《哈姆雷特》汉译文本中显化现象及其动因研究[J],《外语研究》(2):72-80.

[214] 胡显耀.(2005),用语料库研究翻译的普遍性[J],《解放军外国语学院学报》(3):45-48.
胡显耀.(2007),基于语料库的汉语翻译小说词语特征研究[J],《外语教学与研究》(3):214-219.

[215] 胡显耀.(2008),《现代汉语语料库翻译研究》[M],上海:外文出版社.

[216] 黄伯荣、廖序东(主编).(2002),《现代汉语下》(增订第三版)[Z],北京:高等教育出版社.

[217] 黄昌宁.(2002),《语料库语言学》[M],北京:商务印书馆.
[218] 黄立波、王克非.(2006),翻译普遍性研究反思[J],《中国翻译》(5):36-40.
[219] 黄立波、柯飞.(2005),从多个视角探究翻译[J],《外语与翻译》(2):14-17.
[220] 黄立波.(2007),《基于汉英\英汉平行语料库的翻译共性研究》[M],上海:复旦大学出版社.
[221] 黄立波.(2008),英汉翻译中人称代词主语的显化[J],《外语教学与研究》(6):454-459.
[222] 柯飞.(2002a),双语库:翻译研究新途径[J],《外语与外语教学》(9):35-39.
[223] 柯飞.(2002b),语料、网路与口笔译教学[J],《外语教学与研究》(3):231-232.
[224] 柯飞.(2003),汉语把字句特点、分布及英译研究[J],《外语与外语教学》(12):1-5.
[225] 柯飞.(2005),翻译中的隐和显[J],《外语教学与研究》(4):303-307.
[226] 克里斯特尔·戴维(编).(2000),《现代语言学词典》(第四版)(*A Dictionary of Linguistics and Phonetics*),沈家煊译,北京:商务印书馆.
[227] 李德超、王克非.(2009),平行文本比较模式与旅游文本的英译[J],《中国翻译》28(4):54-58.
[228] 李德凤、胡枚.(2006),学习者为中心的翻译课程设置[J],《外国语》(2):59-65.
[229] 连淑能.(1993/2002),《英汉对比研究》[M],北京:高等教育出版社.
[230] 廖七一.(2000),语料库与翻译研究[J],《外语教学与研究》32(5):380-384.
[231] 廖七一.(2001),《当代英国翻译理论》[M],武汉:湖北教育出版社.
[232] 林煌天.(2005),《中国翻译词典》[Z],武汉:湖北教育出版社.
[233] 蔺璜.(1997),现代汉语介词的语法作用[J],《语文研究》(2).
[234] 刘康龙、穆雷.(2006),语料库语言学与翻译研究[J],《中国翻译》(1):59-64.
[235] 刘礼进.(1997),英汉人称代词回指和预指比较研究[J],《外国语》(6):40-44.
[236] 刘宓庆.(2002),《文体与翻译》[M],北京:中国对外翻译出版公司.
[237] 刘宓庆.(2005),《中西翻译思想比较研究》[M],北京:中国对外翻译出版公司.
[238] 刘士聪.(2002),《汉英英汉美文翻译与鉴赏》[M],南京:译林出版社.
[239] 陆俭明、马真.(1985),《现代汉语虚词散论》[M],北京:北京大学出版社.
[240] 吕俊.(2002),翻译学应从解构主义那里学什么——对九十年代中期以来我国译学研究的反思[J],《外国语》(5):48-54.
[241] 吕叔湘(主编).(1980/1999),《现代汉语八百词》(增订本)[M],北京:商务印书馆.
[242] 吕叔湘、朱德熙.(1979),《语法修辞讲话》[M],北京:中国青年出版社.
[243] 罗选民.(2001),《话语分析的英汉语比较研究》[M],长沙:湖南人民出版社.
[244] 马建忠.(1986),《马氏文通读本》[M],上海:上海教育出版社.
[245] 钱军等(译注),(2001),《雅柯布森文集》(罗曼·雅柯布森著)[M],长沙:湖南教育出版社.
[246] 秦洪武、柯飞.(2004),基于语料库的翻译语言分析——以"so... that"的汉语对应结构为例[J],《现代外语》27(1):40-48.
[247] 秦洪武、王克非.(2007),对应语料库在翻译教学中的应用:理论依据和实施原则[J],《中国翻译》(5):43-48.

[248] 秦洪武、王克非.(2004),基于语料库的翻译语言分析——以 so... that 的汉语对应结构为例[J],《现代外语》(1):44-52.
[249] 任晓霏.(2002),从形合意合看汉英翻译的形式对应[J],《中国翻译》(3):33-35.
[250] 邵志洪.(2005),《汉英对比翻译导论》[J],上海:华东理工大学出版社.
[251] 申丹(编).(2008),《西方文体学的新发展》[C],上海:上海外语教育出版社.
[252] 申丹.(1995),《文学文体学与小说翻译》[M],北京:北京大学出版社.
[253] 申丹.(2000),《西方现代文体学百年发展历程》[J],《外语教学与研究》.
[254] 申连云.(2003),形合与意合的语用意义及翻译策略[J],《外国语》(2):67-73.
[255] 思果.(2002),《译道探微》[M],北京:中国对外翻译出版公司.
[256] 宋志平.(2003),英汉语形合意合对比研究综观[J],《东北师大学报(哲社版)》(2):92-98.
[257] 谭载喜.(2005),《翻译学》[M],武汉:湖北教育出版社.
[258] 王灿龙.(2000),人称代词"他"的照应功能研究[J],《中国语文》(3):228-237.
[259] 王力.(1985),《王力文集》(第二卷)[M],济南:山东教育出版社.
[260] 王力.(1989),《中国现代语法》[M],北京:商务印书馆.
[261] 王宇.(2003),关于本科翻译教学的再思考——探索"以学习者为中心"的翻译教学模式的一次尝试[J],《外语界》(1):17-25.
[262] 王克非、秦洪武、王海霞.(2006),双语对应语料库翻译教学平台的应用[J],《外语电化教学》(6):43-49.
[263] 王克非、秦洪武、王海霞.(2007),双语对应语料库翻译教学平台的应用初探[J],《外语电化教学》(6):3-8.
[264] 王克非、胡显耀.(2008),基于语料库的翻译汉语词汇特征研究[J],《中国翻译》(6):16-21.
[265] 王克非、黄立波.(2007),语料库翻译学的几个术语[J],《四川外语学院学报》(6):101-105.
[266] 王克非.(1997),论翻译研究之分类[J],《中国翻译》(1):12.
[267] 王克非.(2003),英汉/汉英语句对应的语料考察[J],《外语教学与研究》(6).
[268] 王克非.(2004a),新型双语语料库的设计与构建.《中国翻译》(6):73-75.
[269] 王克非.(2004b),《双语对应语料库:研制与应用》[C],北京:外语教学与研究出版社.
[270] 王克非.(2004c),双语平行语料库在翻译教学上的用途[J],《外语电化教学》(6):27-32.
[271] 王克非.(2006),语料库翻译学——新研究范式[J],《中国外语》(3):8-9.
[272] 王克非.(2007),双语对应语料库翻译教学平台的应用初探[J],《外语电化教学》(6):55-61.
[273] 王克非、秦洪武.(2009),英译汉语言特征探讨——基于对应语料库的宏观分析[J],《外语学刊》(1):3-8.
[274] 王琼琼、王克非.(2004),一项翻译实验分析和语料库分析——从第三人称代词照应看翻译敏感的培训[J],载王克非等,《双语对应语料库:研制与应用》[M],北京:外语教序

与研究出版社,223-231.
[275] 王宗炎.(2000),《语言问题探索》[M];广州:中山大学出版社.
[276] 韦瑶瑜、雷蕾.(2006),双语平行语料库辅助中国英语学习者翻译实践[J],《基础教育外语教学研究》(4):28-30.
[277] 吴昂、黄立波.(2006),关于翻译共性的研究[J],《外语教学与研究》(5):296-302.
[278] 吴一安.(2005),优秀外语教师专业素质探究[J],《外语教学与研究》(3):199-205.
[279] 谢耀基.(1990),《现代汉语欧化语法概论》[M],台北:光明图书公司.
[280] 谢耀基.(2001),汉语语法欧化综述[J],《语文研究》(1).
[281] 熊学亮.(2007),《语言使用中的推理》[M],上海:上海外语教育出版社.
[282] 徐赳赳.(2003),《现代汉语篇章回指研究》[M],北京:中国社会科学出版社.
[283] 徐盛桓.(2007),基于模型的语用推理[J],《外国语》(3):2-9.
[284] 许国璋.(1991),《许国璋论语言》[C],北京:外语教学与研究出版社.
[285] 许家金.(2004),兰开斯特汉语语料库介绍[OL],http://fld.ustc.edu.cn/123/xujiajin/papers/.
[286] 杨惠中(主编).(2002),《语料库语言学导论》[M],上海:上海外语教育出版社.
[287] 杨惠中.(2004),编者按[J],《现代外语》(2).
[288] 叶常青.(2003),自建语料库在翻译教学中的应用——《红楼梦》中英文本用于翻译教学的课堂设计[J],《外国语言文学》(3):41-44.
[289] 叶莎妮、吕雅娟、黄赟等.(2008),基于 Web 的双语平行句对自动获取[J],《汉语信息学报》(5):67-73.
[290] 余光中.(2002),《余光中谈翻译》[M],北京:中国对外翻译出版公司.
[291] 于连江.(2004),基于语料库的翻译教学研究[J],《外语电化教学》(6):40-44.
[292] 俞士汶、常宝宝、詹卫东.(2003),《计算语言学概论》[M],北京:商务印书馆.
[293] 张虹.(2001),高名凯、吕叔湘、王力四十年代代词理论比较[J],《遵义师范学院学报》,(4):58-61.
[294] 张斌、张谊生.(2000),《现代汉语虚词》[M],上海:华东师范大学出版社.
[295] 张培基等.(1980),《英汉翻译教程》[M],上海:上海外语教育出版社.
[296] 赵宏、邵志洪.(2002),英汉第三人称代词语篇照应功能对比研究[J],《外语教学与研究》(3):174-179.
[297] 赵世开.(1996),The contrastive study of English and Chinese personal pronouns: a preliminary pragmatic analysis(英汉人称代词对比研究)[C],载李瑞华(主编),《英汉语言文化对比研究》[C],上海:上海外语教育出版社,185-199.
[298] 赵世开(主编).(1999),《汉英对比语法论集》[C],上海:上海教育出版社.
[299] 周小兵.(1997),介词的语法性质的介词研究的系统方法[J],《中山大学学报》(3):109-115.
[300] 周有斌.(1999),国内百年汉语代词研究概述[J],《淮北煤炭师院学报》(哲社版),(4):91-95.

后　记

我开始语料库翻译学研究不觉已经十个年头了。起因是2000年申请获得教育部人文社科重点研究基地重大项目的支持，创建了当时最大的通用汉英平行语料库，并开展了一系列初步探讨，结集为《双语对应语料库：研制与应用》，2004年由外语教学与研究出版社出版。翌年，国家社科基金批准了我申请的"基于大型英汉对应语料库的翻译研究与教学平台"，使我的语料库翻译研究得以延续下来。又由于我的一些学生和朋友陆续进入这一研究领域，团队不断扩大，产出不少让人感觉新颖独到的成果，于是，基于语料库的翻译研究在我国逐渐形成一个新的学科分支——语料库翻译学。

这本书，特别是概论部分，已多处论及语料库翻译学的缘起、发展和趋势。这虽然只是一个不到20年的年轻研究领域，但在国际上和国内都日益受到重视。研究论文论著越来越多的涌现，专题学术研讨会每年不止一次的召开和各种项目、课题的申请与获批，就是学科兴盛的表现。

本书研究的问题和涉及的方面，也是语料库翻译学发展的一个缩影。从研究内容上讲，既有对翻译共性假设的探究和检验，也有对语料库方法论的探寻和反思。在对比模式上，我们已不限于单纯的平行或类比，还有多类文本的复合对比，如平行库加类比库或平行库加单语库的翻译语言和中介语的分析。我们还特别注重探讨基于英语和其他欧洲语言的翻译共性研究结果是否适用于汉语等非印欧语系的语言。从方法论上讲，既注重方法论的解释力，包括如何选择研究对象、如何呈现研究发现以及如何解释研究结果，同时，也认识到语料库研究是一个不断检验和完善假设的过程，它是在实证基础上的描写，而描写最终还是为了更合理的阐释翻译现象。从研究视角上讲，主要有对翻译共性问题、翻译语言特征、翻译文体特征、语料库教学平台以及专门用途语料库的研制与应用等课题的探索，是一个多元化的视角，切入点除形式方面的词汇表现、搭配组合外，还延伸到句法结构、话语特点和文体风格等。不过书中的研究基本还是依据文本的语料库，未包括口译语料库。研究也有待深化原有课题，解决更多问题。

本书四大部分、23章，分别概述语料库翻译学的基本理念和主要内容、探析包

括平行语料库和类比语料库基础上的各类翻译问题研究,尝试基于语料库的翻译教学和实验,并对语料库在翻译研究与教学上的应用问题作一些探讨。全书写作历时五年,涉及问题庞杂,内容丰富而有一定争议。即使这样,本书也不能讨论到语料库翻译学的所有内容,前后分析的视角和观点也可能不尽一致。因此本书只是在语料库翻译学方面的一些探索。这是一个新兴的分支学科,还在发展之中,尚有不少理论、途径、方法、工具等方面的问题需要不断地争论和开发,书中不当之处还望读者不吝指正,大家共同探讨,求得新的进步。

本书是项目研究成果,首先应该感谢全国哲学社会科学规划办的支持和鼓励。本书的完成,我要特别感谢秦洪武、胡显耀、黄立波、武光军诸位。他们在攻读博士学位或做博士后研究期间,参与了这个项目的研究,分担了不少研究课题,书中一些章节也是同他们一起撰写的。还有熊文新、李德超、王金铨博士也部分参与了项目研究并共同撰写了若干章节内容,我也在此表示谢意。这是一个研究团队的集体合作成果,否则是不可能完成这么大的项目和做出这么多成绩来的。最后我要表达对北外以文秋芳教授为主任的中国外语教育研究中心这个非常团结合作的科研集体的感谢!良好的科研环境,共同的学术上进,敦促了我们的研究工作,消解了遇到的困难和烦恼,是真正意义上的科研保障。

两年前的一次学术会议上,上海交通大学出版社的领导,特别是同样在关注语料库的管新潮老师以及张国华副编审,敏锐地意识到这一新兴研究的价值,参与策划和筹备了“语料库翻译学文库”。他们为这套丛书包括本书的精心编辑和顺利出版付出了许多心血,令我感佩,在此一并致谢!

作 者

2011 年 11 月于北外